Local Intensity Ordering based Binary Patterns for Image Region Description

[1]Rajkumar Kannan, [2]Suresh Kannaiyan
Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirappalli, India
[1]rajkumar@bhc.edu.in, [2]sureshk.naga@gmail.com

ABSTRACT

Local image region description is a fundamental task for image feature matching in the field of Computer Vision. A good image region descriptor should have the ability to discriminate image features even though the images differ due to photometric variations and geometric transformations. Over these years, many local region descriptors have been proposed to tackle the aforementioned challenges. Achieving rotation invariance in keypoint description is considered one of the main challenges in local region description and matching. Previous approaches proposed to tackle rotation variations depend on unreliable dominant orientation estimation. In this paper, two novel local image region descriptors called Local Intensity Order-based Center Symmetric Local Binary Patterns (LIOCSLBP) and Local Intensity Order-based Orthogonally Combined Local Binary Patterns (LIOOCLBP) are proposed to build rotation invariant local region descriptions. The rotation invariance characteristic of the proposed binary pattern-based local region description is achieved by applying a simple and efficient mechanism called Local Intensity Ordering (LIO). The proposed descriptors use double interest regions for each interest point to improve feature discrimination. In order to further improve the feature discrimination ability RGBLIOCSLBP, RGBLIOOCLBP, HSVLIOCSLBP and HSVLIOOCLBP are also proposed exploiting RGB and HSV color models. Extensive experiments are conducted to evaluate the performance of the proposed descriptors on standard benchmark datasets for image matching, object recognition and scene recognition against the state-of-the-art descriptors. The experimental results show that the proposed descriptors are highly competitive to several state-of-the-art local region descriptors where the proposed descriptors outperformed the comparative approaches in many cases.

Keywords: Image region descriptor, image feature matching, local binary pattern, local intensity ordering, object recognition, texture classification

1   Introduction

Finding correspondences between local features in two images containing similar or same visual scenario is a core and interesting challenge in Computer Vision. Keypoint description and matching has been used in many computer vision applications including object recognition [1], object classification [2] and image stitching for panorama generation [3]. The keypoint descriptor used in these applications needs to be invariant to various geometric and photometric image transformations such as scale, rotation, viewpoint, illumination and blur. The main aim of designing an invariant local image region descriptor is to build discriminative features while also maintaining robustness [4] to various geometric and photometric image transformations. A fundamental approach to find the image feature correspondences comprises three steps [4]: 1) keypoints or interest points detection,

2) feature description, and 3) feature matching. Each of these steps is challenging whereby, mainly the feature description needs to be both discriminative and invariant to various geometric and photometric transformations.

Numerous keypoint descriptors [1, 5, 6, 7, 8, 9, 10] have been proposed over these years to handle the aforementioned challenges in keypoint description. Invariance description to keypoints appear in different scales and is often achieved by estimating the characteristic scales of keypoints in an image [11]. The description approach proposed in [12] automatically selects scale by detecting the keypoints using multi-scale representation to make feature description invariant to scale changes across images. Lowe [1] proposed the idea of utilizing Difference-of-Gaussian (DoG) mechanism to detect scale invariant keypoints. Since the keypoint detection approach used in these descriptors is not designed to adapt to affine changes, the performance of these methods degrade when the images have viewpoint variations. Harris-Affine [13] and Hessian-Affine [14] are the two most widely used affine-invariant keypoint detection approaches for dealing with viewpoint variations. These affine-invariant keypoint detection approaches detect local image regions which are covariant to affine transformations. This is achieved by estimating the shape of elliptical image regions and normalizing them into circular image regions. Most of the existing descriptors rely upon estimating the dominant orientation of an image patch to achieve rotation invariance where each interest region is rotated according to its dominant orientation [1]. However, these descriptors often fail to build a rotation-invariant feature description, which further leads to mismatch among feature correspondence due to error in estimating the dominant orientation [4]. A few descriptors have also been proposed to achieve illumination invariant feature description [15, 16].

Most of the earlier keypoint descriptors use gradient orientation histograms for keypoint description. For applications such as object recognition and image retrieval, computation speed is expected to be high in cases where the slower speed of gradient orientation histograms computation makes it unsuitable for real-time performance. This has inspired researchers to further propose binary operators based keypoint description approaches where the use of binary operator makes the local region description simple and fast. Among them, Local Binary Pattern (LBP) [17] is one the earliest and most successful binary descriptors proposed for computing binary patterns for feature description. The feature description based on binary pattern estimation is comparatively much faster than the gradient based keypoint description schemes. The original LBP [17] produces 256-dimensional histogram for 8-neighboring pixels, which makes the keypoint matching computationally little expensive. In order to reduce the computational overhead experienced in keypoint matching caused by the higher dimensionality of Local Binary Pattern (LBP) [17] based image region descriptors, Center-symmetric LBP (CSLBP) [15] and Orthogonal Combination of LBP (OCLBP) [18] are proposed. The CSLBP [15] is proposed for reducing the dimension of LBP features by comparing a pixel with its center symmetric pixel pairs instead of comparing each pixel with the center pixel. The OCLBP [18] method reduces the LBP dimension by computing binary patterns using orthogonal combination of neighboring pixels. Even though the CSLBP and OCLBP reduce the LBP dimension and computation complexity, they still construct discriminative feature description.

Albeit the binary descriptors are compact, faster and easy to compute, they still depend on the dominant orientation estimation to build the keypoint description invariant to rotation changes. This motivates us to build a good binary pattern description approach that consistently performs better in achieving rotation invariance. The core aim of this paper is to achieve rotation invariance characteristic in computing the binary pattern itself rather than incorporating unreliable and

computationally expensive dominant orientation estimation for rotation invariant local region description. Recently, Local Intensity Ordering (LIO) is gaining attention for extracting rotation and illumination invariant [16] features in applications such as image matching [16] and face recognition [19]. In this paper, a novel methodology for binary pattern based interest region description based on LIO is proposed. In contrast to [16] which used LIO for interest region segmentation and [19] which used LIO for face representation, in our paper, the LIO is used for local neighborhood pixel sorting to make the binary pattern computation invariant to rotation changes. Two binary pattern based local region descriptors are designed based on the CSLBP and the OCLBP. Hence, two different features are built, namely Local Intensity Order based Center Symmetric Local Binary Patterns (LIOCSLBP) and Local Intensity Order based Orthogonally Combined Local Binary Patterns (LIOCSLBP). It should be noted that the each of the names LIOCSLBP and LIOOCLBP is used to refer both the binary operator and the corresponding region descriptor interchangeably.

The main contributions of this paper are as follows:

1. Two novel texture operators called LIOCSLBP and LIOOCLBP are proposed using the Local Intensity Ordering (LIO). The LIOCSLBP and LIOOCLBP are the rotation invariant extension of the popular CSLBP and OCLBP texture operators respectively. The rotation invariance is achieved in computing the binary pattern itself rather than exploiting expensive dominant orientation based region description mechanism.

2. Two local image region descriptors using the LIOCSLBP and LIOOCLBP operators are proposed. The proposed descriptors produce less dimensional features than the original LBP and generate robust features for rotation variations. Since LIO is always stable and invariant to rotation changes, the binary pattern computed after applying LIO is also invariant to rotation changes and reliable than the descriptors that use reference orientation to achieve rotation invariance.

3. Two different patches may have locally similar appearance that may lead to ambiguity when matching features between two images. Use of multisupport region for image patch description can solve this issue. Hence to improve the discrimination ability, the proposed descriptors use two support regions around each interest point. This makes the features computed by the proposed descriptors highly discriminative to cope with ambiguous matching problems.

4. Four novel color descriptors based on LIOCSLBP and LIOOCLBP adapting RGB and HSV color models are also proposed. Unlike the traditional keypoint descriptors that use gray image patch for feature description, the proposed color descriptors compute features from RGB and HSV color models. Incorporation of color information to feature description makes the features invariant to monotonic illumination changes while also keeping invariance to rotation change. This also helps the feature description scheme to discriminate two different patches with similar gray scale values which preserves one-to-one keypoint matching.

The remainder of this paper is organized as follows: In section 2, previous work on image descriptors is discussed. In section 3, a detailed review on LBP, RLBP, CSLBP and OCLBP is presented. The details of the local image patch descriptions using the LIOCSLBP and LIOOCLBP operators are presented in section 4, where the improvements of the LIOCSLBP and LIOOCLBP with color information are also presented. The experimental results for texture classification, image matching and object recognition are reported in section 5. Finally, conclusion and future work of this paper are presented in section 6.

2   Related work

This section reviews the literature on local image descriptors and bag-of-visual-words representation based object category recognition. The existing work discussed in the section are categorized into image region descriptors, texture operators, color adapted image region descriptors and bag-of-visual-words.

2.1   Image region descriptors

Keypoint detection and keypoint feature description are the two important steps in constructing local feature description of an image. The feature detector detects feature points (i.e. keypoints, interest points or regions) in an image. The region around the interest point is taken for feature description that should have the ability to distinguish features of one image to another even though there are variations in scale, rotation, blur, illumination and view point. To achieve this, many keypoint description schemes have been proposed in the past decades. Among them, Scale Invariant Feature Transform (SIFT) [1] has been considered as a standard and one of the most successful descriptors for image matching. The SIFT descriptor achieves scale invariance using Difference of Gaussian (DoG) scheme where the rotation invariance is achieved by estimating dominant orientation of the region. It produces 128-dimensional histogram of gradient orientations and locations.

Inspired by the SIFT descriptor, many keypoint descriptors have been proposed over these years. The PCA-SIFT [5] used Principal Component Analysis (PCA) to construct a compact and discriminative descriptor. Although PCA-SIFT reduces SIFT descriptor's dimensions, it generates less distinctive features than the SIFT descriptor's features. Gradient Location and Orientation Histogram (GLOH) [6] is a modified version of the SIFT, which exploits log-polar grid to split interest region instead of the Cartesian location grid used in the SIFT. In addition, it also exploits PCA to reduce the dimensionality of the descriptor. SURF descriptor [7] uses Hessian matrix for interest point detection where Haar-Wavelet response and integral images are utilized for region description. Incorporation of integral images for region description in SURF achieves 3 to 7-fold speed-ups than the SIFT descriptor. However, SURF is sensitive to more complicated geometric deformations such as large view point and rotation changes [20]. Hence, the MDGHM-SURF is proposed using modified discrete Gaussian–Hermite moment (MDGHM), which devises a movable mask to represent the local feature information of non-square images. The MDGHM preserves more feature details than the SIFT and the original SURF descriptor [20].

Berg and Malik [21] proposed a keypoint descriptor to handle geometric blur in image matching. A few methods used Weber's law for illumination invariant feature description [22, 23]. Chen *et al*. [22] proposed a descriptor named WLD using Weber's law to improve feature discrimination under varying lighting conditions. Their approach is based on the fact that human perception of a pattern depends on not only the change of a stimulus (such as sound, light, etc.) but also the original intensity of the stimulus [22]. It uses differential excitation and gradient orientation to construct a concatenated WLD histogram feature for a given image. Winder and Brown [24] proposed local descriptors using different combinations of local features and feature pooling schemes. Mittal and Ramesh [25] proposed Intensity-Augmented Ordinal Measure to find visual correspondence. They developed a new keypoint matching strategy that combines intensity and rank information. Their approach improves the feature robustness to challenges such as Gaussian noise and image compression. Gupta and Mittal [26] proposed a descriptor to overcome illumination variations by utilizing intensity orders. Tang *et al*. [27] used a 2D histogram of positions and intensity orders to overcome illumination problems. Wang *et al*. [16] proposed the Local Intensity Order Patterns (LIOP)

that divides the local patch into subregions using order of intensity. It uses both local and overall intensity ordinal information of the local patch [16], where it produces a highly discriminative descriptor. The CLRF descriptor [28] was proposed using log polar transformation and 2D discrete wavelet transformation to produce a compact descriptor. Most of the aforementioned descriptors used only single interest region around each interest point for description. Use of a single region around each interest region is not always suitable for image description since two non-corresponding local regions accidently may have similar appearance that leads to ambiguity in image feature matching. Thus the keypoint description with single region is not always suitable to find corresponding regions.

The aforementioned issue can be overcome with the use of multiple regions for image region description. Multisupport Region Order-Based Gradient Histogram (MROGH) [4] and Multisupport Region Rotation and Intensity Monotonic Invariant Descriptor (MRRID) [4], interleaved order based local descriptor (IOLD) [29], feature descriptor using entropy rate (FDER) [30] and monotonic invariant intensity descriptor (MIID) [31] are a few local image descriptors which use multisupport regions for image description. The use of multisupport regions improves feature discrimination and increases the dimensionality of the final descriptor where the feature matching becomes computationally little expensive. However recent approaches prefer to use multisupport regions for region description to improve the keypoint matching accuracy.

Most of the SIFT-based descriptors used only histogram of first order gradients for feature extraction. In contrast, HSOG [32] used the second-order gradients to compute the Histogram of Second Order Gradients. The HSOG [32] captures the curvature related geometric properties of the neural landscape i.e. cliffs, ridges, summits, valleys, basins, etc. A feature description approach using patch intensity permutation and zone division is proposed in [33] that improves robustness and invariance to rotation changes and monotonic brightness changes. Spatial co-occurrence of local intensity order [19] is presented to preserve invariance to illumination and to improve feature discriminative power by capturing the correlation between locally adjacent regions. Bosch *et al.* [34] proposed a hybrid descriptor using a generative and discriminative approach for feature description and applied them for scene classification. Recently, texture operators [15, 17, 18] have gained much attention in image region description, because of their computation simplicity and invariance to monotonic illumination changes property. The descriptor approaches related to Local Binary Patterns are given in the following section.

2.2   Texture operators

Local Binary Patterns (LBP) [17] is a powerful technique for texture description that is used in many applications such as local region description, face recognition [19], medical image classification and retrieval [35, 36]. The LBP gained much popularity because of its computational simplicity in encoding textured images and tolerance to monotonic illumination variations. However, the main shortcoming of the LBP is that it produces a rather high dimensional histogram of binary codes and is not too robust for flat images. Following the LBP, many texture operators [15, 37, 38, 39] have been proposed for texture representation and classification. Zhao *et al.* [39] proposed Completed Local Binary Count (CLBC) for rotation invariant texture classification. The performance of the aforementioned texture operators in texture classification inspires to exploit them for image region description.

Recently, LBP based texture operators [15, 18] are utilized for local image patch description. Their main objective is to reduce the dimension of the histogram feature of original LBP and utilize binary patterns for region description. Heikkila *et al*. [15] proposed the CSLBP descriptor for reducing LBP

feature dimension and improving the description robustness for flat images. The CSLBP compares center symmetric pairs of pixels instead of comparing neighbors with central pixel, which reduces half computation of the LBP. The Orthogonal Combination of LBP (OCLBP) [18] is proposed to reduce the dimension of the LBP and to improve LBP descriptor's discrimination ability. Similar to binary pattern descriptors, ternary pattern descriptors are also proposed [40] to improve feature discrimination. However, the ternary pattern descriptor construction is computationally intensive than the binary pattern computation. Qi *et al.* [41] proposed Pairwise Rotation Invariant Co-Occurrence Local Binary Pattern operator which used spatial co-occurrence and orientation co-occurrence to improve discrimination and rotation invariance. Yang and Cheng [42] proposed a feature descriptor using local difference binary that is fast and distinctive [42]. Recently, local diagonal extrema pattern [36], local derivative quantized binary pattern [43] are proposed for CT image retrieval and object recognition respectively. Li *et al.* [44] proposed Dominant Center-Symmetric Local Binary Pattern (DCSLBP) for improving the CSLBP descriptor. The DCSLBP [44] produces half dimensional feature descriptions of the original CSLBP features. Such lower dimensionality makes it to lose the feature discrimination ability in image matching. Although the aforementioned LBP based descriptors achieved impressive image matching and image classification accuracy, those approaches are still sensitive to rotation changes since they depend on dominant orientation estimation. In this work, we aim to achieve rotation invariant binary pattern computation replacing the dominant orientation based patch normalization approach.

2.3 Color adapted image region descriptors

Though most of the feature descriptor methods utilize gray image patches, color adapted feature descriptors have also gained considerable attention in the recent years. Many SIFT based color descriptors have been proposed [34, 45, 46, 47] which improve their discriminative power using color models such as RGB and HSV. Abdel-Hakim and Farag [45] proposed CSIFT descriptor utilizing color information. Van de Weijer *et al.* [46] proposed a descriptor using statistics of color image derivatives. Van de Sande *et al.* [48] discovered that the color SIFT descriptors outperforms the original SIFT descriptor for object and scene recognition. Recently color adapted LBP based descriptors [18] are proposed to improve feature discrimination and their performances are evaluated for Bag-of-Words (BoW) based object recognition. For a detailed evaluation of different color descriptors, the readers are recommended to refer Burghouts and Geusebroek [47]. Despite the fact that computing region description for each color channel increases the dimensionality of the final descriptor, the performance of these color descriptors suggests that the region description becomes very discriminative and invariant to illumination changes. Thus, the color information is utilized in this work, where four color descriptors are proposed based on the RGB and HSV color models.

2.4 Bag-of-Visual-Words (BoVW)

Bag-of-Words (BoW) modeling is a text quantization method which was initially used for text retrieval. Its simplistic nature and efficient histogram representation inspired researchers to use BoW for image representation and image retrieval [49]. Following [50], many Bag of Visual Words (BoVW) based image representation methods [51, 52] have been proposed. Vedaldi *et al.* [53] used Multiple Kernel Learning [49] technique for image representation that significantly improved performance in image classification. Using spatial information, Lazebnik [54] proposed Spatial Pyramid Matching Scheme for image feature matching. Fisher vectors [55], VLAD [56] and super-vectors [57] have been put forward, all utilizing the difference between the local descriptor and each code word that improves BOW coding scheme to vectorial coding scheme. Though there are many image content

representation techniques, BOVW is highly used for image representation [18] because of its effectiveness in object recognition applications.

3   Review on LBP, RLBP, CSLBP and OCLBP

This section presents a detailed review on the state-of-the-art texture descriptors LBP, RLBP, CSLBP and OCLBP, and explains the computation of these descriptors using examples.

3.1   Local Binary Patterns

Ojala *et al.* [17] introduced the local binary pattern operator for texture analysis where the histogram of binary patterns computed over a region is used for texture description [17]. Fig.1 illustrates the computation of LBP code for a 3×3 image patch. For each center pixel n_c in a region/image the binary pattern code $LBP_{R,N}$ is computed using Equation (1). Firstly, a binary value for each surrounding pixel is computed based on two conditions. If the gray value of the neighboring pixel n_i is greater than or equal to the gray value of the center pixel n_c, then the binary value is set to 1, otherwise 0. Binary values of the neighboring pixels are then multiplied with their corresponding positional weights where the resultant values are summed to produce the final LBP value. The equation for computing LBP of a pixel is,

$$LBP_{R,N} = \sum_{i=0}^{N-1} s(n_i - n_c)2^i, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

where n_c represents the gray value of the center pixel, N is the number of neighboring pixels, x is the difference between center pixel n_c and neighboring pixel n_i . The value n_i ($i = 0,…,N-1$) denotes the gray value of a neighboring pixel on a circle of radius R. Bilinear interpolation can be used to get neighboring pixel value, when the neighboring pixels do not fall in the center of pixels [21]. The radius R is usually assigned to a small value, since the correlation between pixels decreases when the distance increases. A histogram is computed over each interest region using LBP value of each pixel. This binary pattern based histogram representation is highly discriminative, robust against illumination changes, easy to compute and does not require many parameters to be set [22]. Thus, LBP is considered as a good candidate preferred for local image description. For given a 3×3 neighboring pixels, LBP will produce 2^8 dimensional histogram features.

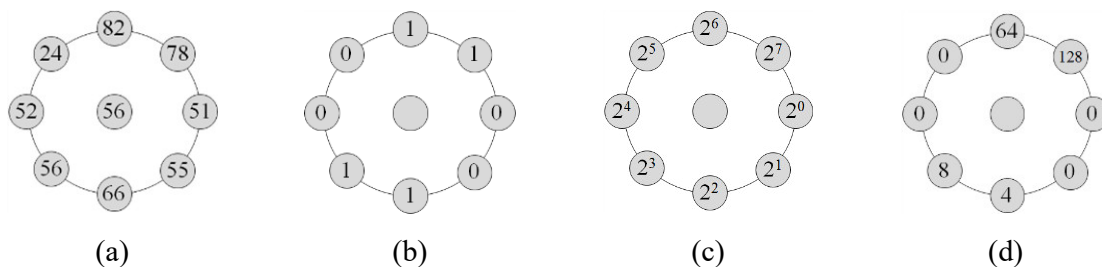(a)   (b)   (c)   (d)

Figure. 1. Computation of the Original LBP code. (a) 3×3 gray patch. (b) Thresholded neighbors of (a). (c) Weights for (b). (d) Multiplication of (b) and (c). The LBP code for the above given patch is 4+8+64+128 = 204.

3.2   Robust Local Binary Pattern (RLBP)

Yang, *et al.* [58] introduced the Robust Local Binary Pattern (RLBP) using Average Local Gray level (ALG) of each patch for binary pattern computation. The RLBP uses ALG as central pixel value to build features that are insensitive to noise and invariant to monotonic gray scale transformation [58]. The ALG computation for a 3×3 patch is computed as:

$$\mathrm{ALG} \; = \; \frac{n_c + \sum\limits_{i=1}^{8} n_i}{9} \tag{2}$$

where n_c is gray intensity of the center pixel and n_i is the set of gray values of the neighboring pixels. The descriptor which is built using the ALG claimed to be more robust to noise than the descriptor that uses gray value of the central pixel [58]. The illustration of RLBP computation is given in Fig. 2.
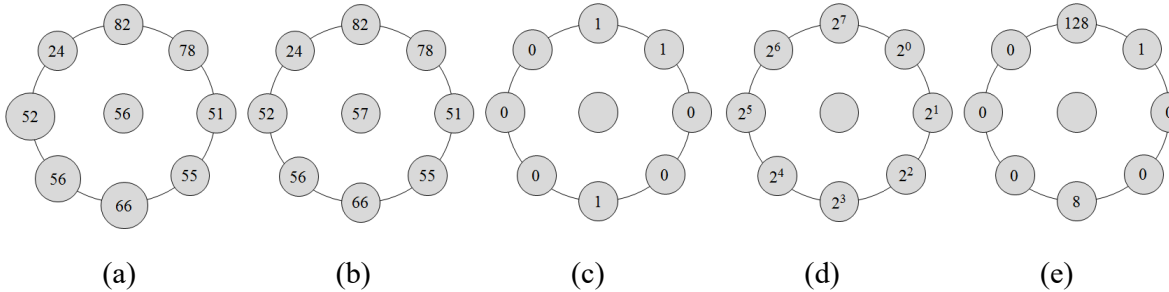

| (a) | (b) | (c) | (d) | (e) |

Figure. 2. RLBP binary pattern computation for a 3×3 patch. (a) A 3×3 patch. (b) Patch with ALG. (c) Thresholded neighbors of (b). (d) Weights for (b). (e) Multiplication of (c) and (d). The RLBP code for the above shown patch is 1+8+128 = 137

The binary pattern computation of RLBP is the same as LBP. The main difference is that the RLBP uses ALG as a threshold to compute binary codes instead of gray value of the center pixel. The RLBP is computed as:

$$RLBP_{R,N} = \sum_{i=0}^{N-1} s(n_i - ALG)2^i, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

where n_i (i = 0,1,..,7) represents the gray values of the neighbor pixels on a circle of radius R. Since the ALG is used as a threshold to compute descriptor, the RLBP is obviously insensitive to noise distortion [58]. Sometimes, the LBP produces the same binary code for different patches whereas the use of the ALG to compute LBP overcomes this issue. The dimension of histogram produced by RLBP is the same as that of LBP. Generally, the histograms produced by both LBP and RLBP is considered high dimensional. Such high dimensional features are highly discriminative but less generalizable and slow in texture classification. Hence, an approach to reduce the histogram dimension of the LBP is needed.

3.3 Center Symmetric Local Binary Patterns (CSLBP)

The LBP descriptor produces a high dimensional histogram which make the feature matching very time consuming. Inspired by the LBP, Marko Heikkil *et al*. [22] proposed the compact version of the LBP called CSLBP that reduces the dimension of LBP histogram. An illustration of the CSLBP computation is shown in Fig. 3. In contrast to LBP which compares each surrounding pixel with the center pixel, the CSLBP ignores the center pixel where the comparison is performed between the center symmetric pixel pairs. The CSLBP is computed as:

$$CSLBP_{R,N,T}(x,y) = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i, \quad s(x) = \begin{cases} 1 & x > T \\ 0 & otherwise \end{cases} \tag{4}$$

where the binary value $s(x)$ is set to 1 if the difference between center symmetric pixels is greater than a threshold T which usually set as 5. For a 3×3 patch, the LBP uses 8 neighbors and 8 comparisons to produce 256-dimensional binary patterns, whereas the CSLBP uses 8 neighbors and only 4 comparisons to produce 16-dimensional binary patterns. The CBLBP descriptor reduces

computation of the LBP by half and achieves robustness on flat image regions [22] and is highly preferred for local image description.
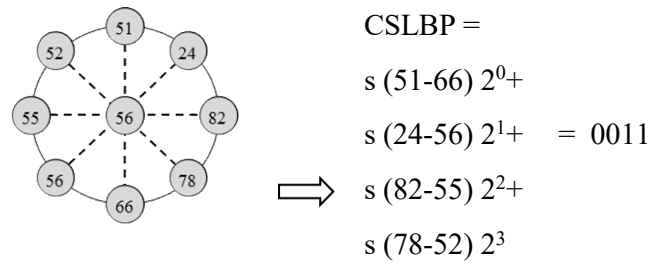
$$CSLBP =$$

$$s\,(51\text{-}66)\,2^0+$$

$$s\,(24\text{-}56)\,2^1+ \quad = 0011$$

$$s\,(82\text{-}55)\,2^2+$$

$$s\,(78\text{-}52)\,2^3$$

Figure. 3. CSLBP computation for a 3×3 patch with threshold *T* = 5

3.4 Orthogonally Combined Local Binary Pattern (OCLBP)

Zhu *et al*. [18] proposed the OCLBP which also aims to reduce the dimension of the original LBP histogram. A sample OCLBP computation for a 3×3 patch is depicted in Fig. 4. The OCLBP uses orthogonal pixels combination to compute LBP histograms. The original LBP operation is first split into two non-overlapping orthogonal groups of pixels. Then the LBP is computed for each orthogonal group separately. The resultant LBPs are concatenated and kept as the OCLBP code. This split-and-merge mechanism reduces the dimensionality of the original LBP histogram. It produces $2^4 \times N/4$ or $4 \times N$-dimensional feature descriptor. For $N = 8$ and $R = 1$, the dimension of the OCLBP features is 32. This clearly shows that the OCLBP is very compact than the original LBP descriptor.
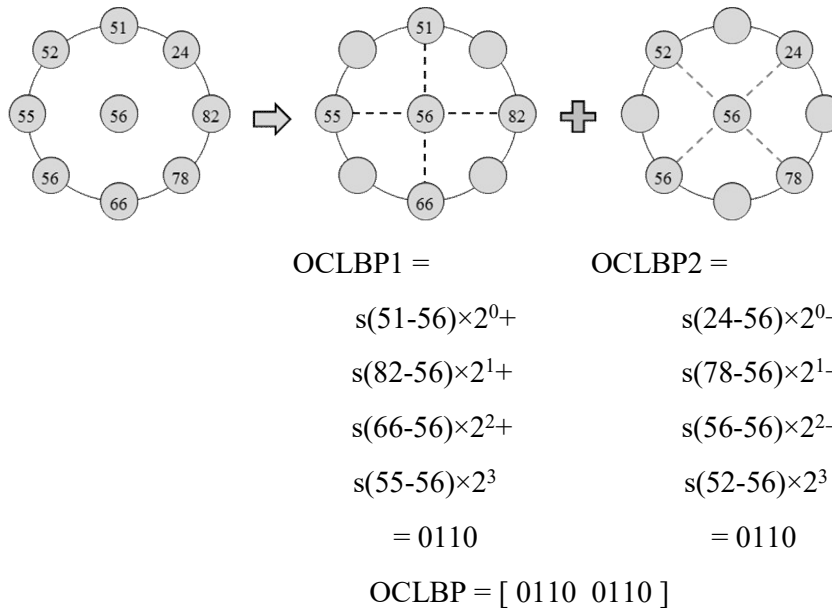
$$OCLBP1 = \qquad\qquad OCLBP2 =$$

$$s(51\text{-}56)\times 2^0+ \qquad\qquad s(24\text{-}56)\times 2^0+$$

$$s(82\text{-}56)\times 2^1+ \qquad\qquad s(78\text{-}56)\times 2^1+$$

$$s(66\text{-}56)\times 2^2+ \qquad\qquad s(56\text{-}56)\times 2^2+$$

$$s(55\text{-}56)\times 2^3 \qquad\qquad s(52\text{-}56)\times 2^3$$

$$= 0110 \qquad\qquad\qquad = 0110$$

$$OCLBP = [\ 0110\ \ 0110\ ]$$

Figure. 4. Illustration of OCLBP computation for a 3×3 patch, where OCLBP1 is computed using vertical and horizontal pixels and OCLBP2 is computed using diagonal pixels. The OCLBP is the concatenation of the OCLBP1 and OCLBP2.

4   The Proposed Method

The previous section has shown a detailed review on binary operators LBP, RLBP, CSLBP and OCLBP. It was evident that the highly efficient texture operators CSLBP and OCLBP aim at reducing the computational overhead experienced in the high dimensional LBP histograms. However, none of the

previous texture operators have incorporated a rotation invariant binary pattern computation yet. In this section, we introduce our novel approach for rotation invariant binary pattern computation.

The key idea of the proposed method is to use intensity ordering of pixel intensities of local patch to build rotation invariant features. Instead of computing dominant orientation of each interest point to make the local feature invariant to rotation, the binary pattern based descriptors exploiting the LIO are computed here. This section presents two novel local feature descriptors called LIOCSLBP and LIOOCLBP improving the CSLBP and OCLBP descriptors respectively. In addition, color information of an image is also exploited for interest region description where four novel color descriptors are also presented in this section. Interest region detection, region partition and region description are the basic steps involving in the local image patch descriptor construction that are explained in the following sections.

4.1 Affine normalized interest region detection

The affine invariant interest points are detected from an image using the Hessian-Affine detector [14] by following the previous local image region descriptors [4, 5, 6]. The interest regions detected in images are shaped either circular or elliptical based on the region detector. To achieve scale or affine invariance, the detected region is usually normalized to a canonical region [4] as shown in Fig. 5. From that a circular region of radius 20.5 pixels is chosen for description for each interest point. Therefore, the region contains 41×41 pixels similar to the descriptors proposed in [4, 5, 6]. Gaussian smoothing is applied to the image when the interest region is larger than the normalized region size. The size ratio of the interest region and the normalized region is used as the standard deviation for Gaussian smoothing as in [4, 6].
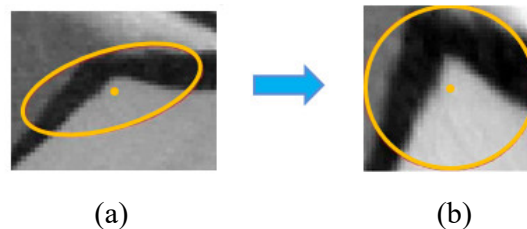

(a)   (b)

Figure. 5. Affine region transformation to circular region [4]. (a). Detected interest point and its affine region. (b). Normalized interest region of (a)

4.2 Interest region partition

The detected interest region can be divided into several ring shaped partitions, where the features computed from each partition can be aggregated together as in RIFT descriptor [59]. However, the features computed from ring based partitioning lose some spatial information, which eventually reduces distinctiveness of keypoint features. Therefore, most of the descriptors [1, 8, 15, 18] divide an interest region into sub-regions to incorporate more spatial information. The proposed descriptors divide the normalized interest regions into *m×m* squared sub-regions (m=2) to preserve spatial information of regions to improve feature discrimination.

4.3 Rotation invariant binary pattern computation

Most of the existing descriptors compute features using either ring shape partitions [59] or dominant orientation [1, 8, 15] to achieve rotation invariance. However, these descriptors are not much invariant to rotation changes where the features based on dominant orientation estimation are not stable enough to achieve effective rotation invariance [4]. In this paper, the Local Intensity Ordering

(LIO) mechanism is exploited to achieve rotation invariance. The LIO of neighboring pixels for a pixel is obtained by sorting the gray scale values of the neighboring pixels into descending order. Consider R is radius, I is intensity, n_c center pixel value and its N neighbors are $\{n_0, n_1,..., n_{N-1}\}$. The LIO at a pixel n_c is the permutation (σ) that sorts the neighboring pixels in decreasing order as:

$$I\left(n_{\sigma(0)}\right)_R \geq I\left(n_{\sigma(1)}\right)_R \geq .... \geq I\left(n_{\sigma(N-1)}\right)_R \tag{5}$$

4.3.1 LIOCSLBP features

LIOCSLBP feature is constructed by computing the CSLBP after applying LIO to neighboring pixels. The LIOCSLBP is computed as:

$$LIOCSLBP_{R,N,T}(x,y) = \sum_{i=0}^{(N/2)-1} s\left(n_i - n_{i+(N/2)}\right)2^i, \quad s(X) = \begin{cases} 1 & X > T \\ 0 & otherwise \end{cases} \tag{6}$$

where (x, y) denotes co-ordinate points of a pixel, n_i and $n_{i+(N/2)}$ correspond to the gray values of center-symmetric pixel pairs of N equally spaced pixels on a circle of radius R. Parameter X is the difference between center symmetric pixel pairs where T is a threshold which is usually set to 1% of the pixel value range. If the pixel intensities are in the range 0 to 1, then T will be usually set to 0.01 [15]. The resultant feature is called LIOCSLBP feature. Sample computation of LIOCSLBP is depicted in Fig. 6. For a 3×3 patch, the LIOCSLBP picks the pixel pairs $(n_0 - n_4)$, $(n_1 - n_5)$, $(n_2 - n_6)$ and $(n_3 - n_7)$ to compute binary pattern. Therefore, it needs only four comparisons to compute the binary patterns for the 3×3 patch. Thus, the LIOCSLBP produces $2^{N/2}$-dimensional histogram features.

4.3.2 LIOOCLBP features

Here, the LIOOCLBP is computed using the orthogonal combination of local binary patterns (OCLBP) after applying the LIO for each pixel in the interest region. The OCLBP based features are extracted using orthogonal combination of local binary patterns that have a lower-dimension than the LBP histogram while keeping more discriminative ability [18]. It is very compact comparing to the LBP features since it produces $2^4 \times (N/4)$ dimensional features. Equation (7) shows the computation of the LIOOCLBP for a pixel with 8 neighboring pixels. The LIOOCLBP is computed as:

$$LIOOCLBP_{R,N=8,T}(x,y) = \sum_{i=0}^{(N/2)-1} s\left(n_{2i} - ALG\right)2^i \parallel \sum_{i=0}^{(N/2)-1} s\left(n_{2i+1} - ALG\right)2^i, \quad s(X) = \begin{cases} 1 & X > T \\ 0 & otherwise \end{cases} \tag{7}$$

where ALG is Average Local Gray level which is computed as $ALG = \dfrac{n_c + \sum_{i=1}^{8} n_i}{9}$ and || denotes the concatenation. The sample LIOOCLBP computation for a pixel and its eight neighboring pixels is given in Fig. 6. The LIOOCLBP first considers horizontal and vertical pixels to compute binary patterns and then it computes binary pattern for diagonal pixels. In each step, it results in 16-dimensional histogram features and the final LIOOCLBP histogram is of 32 dimensions when 8 neighboring pixels are considered. From Fig. 6, it can be noted that the proposed descriptors generate the same binary patterns for both regular and rotated patches whereby the CSLBP and OCLBP approaches generated different patterns.
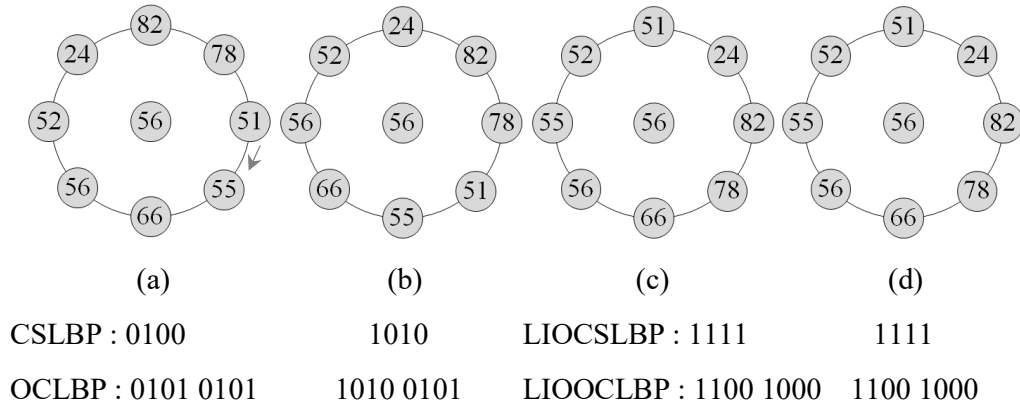
(a)	(b)	(c)	(d)
CSLBP : 0100	1010	LIOCSLBP : 1111	1111
OCLBP : 0101 0101	1010 0101	LIOOCLBP : 1100 1000	1100 1000

Figure. 6. Computation of CSLBP, OCLBP, LIOCSLBP and LIOOCLBP for a 3×3 patch. (a) Original patch (the arrow mark in (a) indicates the starting position to compute binary patterns). (b) Rotated patch, (c) LIO of the original patch and (d) LIO of the rotated patch. (c) and (d) are the resultant binary patterns computed after applying LIO. (The CSLBP and LIOCSLBP are computed by setting threshold T = 5, and LIOOCLBP is computed by setting ALG = 57.78 as center pixel value).

4.4   Local Image Region Descriptor construction

The construction of the local interesting region description using LIO based operators is presented in this section. The steps in building the proposed descriptors are illustrated in Fig. 7, and details of each step are given below.
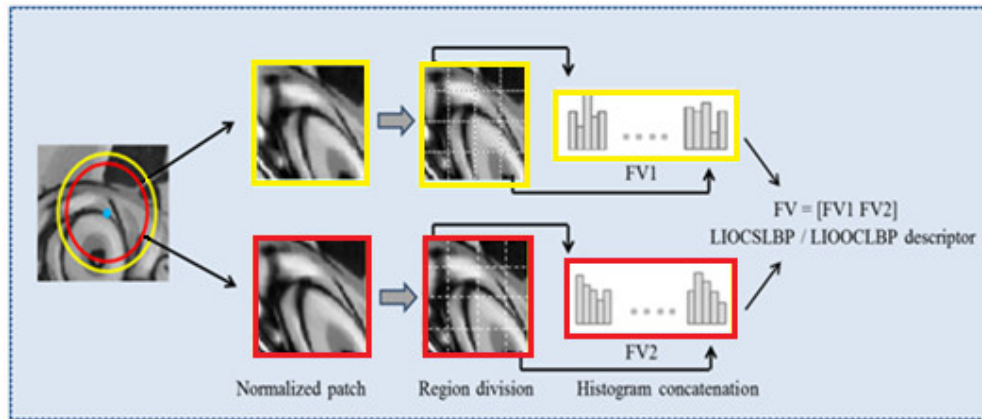

Figure. 7. Block diagram of the LIOCSLBP and LIOOCLBP descriptors construction. Here, FV1 and FV2 are descriptors computed from two patches and FV is the concatenation of them.

The proposed LIOCSLBP and LIOOCLBP operators use each partitioned interesting point region (sub-region) for feature extraction. The histogram features constructed for each sub-region are then concatenated to form the final feature descriptor. Two non-corresponding interest points may have similar appearances in some local region [4]. Thus, the image descriptors constructed using only single support region are not feasible sometimes to distinguish mismatches from the correct ones. To overcome this issue, multiple interest regions are used to construct descriptor to improve discrimination in [4]. The descriptors computed from multiple interest regions are discriminative, but are computationally a little expensive.

In order to balance the trade-off between the discrimination and computational efficiency, two interest regions (R=2) around each interest point are used for feature description. The final descriptor is the concatenation of features computed over two interest regions where its dimension depend on

the parameters such as number of neighboring pixels (N), number of sub-regions of a patch (m) and the number of interest regions (R). Thus, it produces $m \times m \times 2^{N/2}$ and $m \times m \times (2^4 \times N/4)$ dimensional feature descriptors for LIOCSLBP and LIOOCLBP respectively for single interest region. For $m = 2$ and $N = 8$, the dimensions of LIOCSLBP and LIOOCLBP are 64 and 128 respectively. These histogram features are then normalized to unit length. To eliminate the domination of very large histogram elements, a threshold is set to limit all descriptor elements to be less than or equal to 0.2 as done in [18]. After the thresholding, the feature descriptor is renormalized to unit length and kept as final feature description of an interest point.

4.5 Color Descriptors

The proposed descriptors presented in the previous subsections use only gray images for feature description. However, color information of images can enhance the discrimination power of keypoint features [18]. Adding color information to feature description will improve photometric invariance (illumination invariance) that will be useful when images are captured under different lighting conditions. Following the OCLBP [18], the color enhanced descriptors such as RGBLIOCSLBP, HSVLIOCSLBP, RGBLIOOCLBP and HSVLIOOCLBP are proposed using RGB or HSV color models.

The interest points are detected in gray image and their color channels are used only during the region description. In order to embed the color information into the feature description, binary features are extracted from each color channel of the RGB or HSV. Then, the features extracted from each color channel are concatenated to build a color enhanced descriptor. The dimension of color descriptor is 3 times larger than the descriptors (LIOCSLBP and LIOOCLBP) which are built using the gray images. The construction of color descriptor using RGB model is depicted in Fig. 8. Each color space has its own characteristics and advantages, which results in different image matching and object recognition accuracies that are discussed in experimental results section.
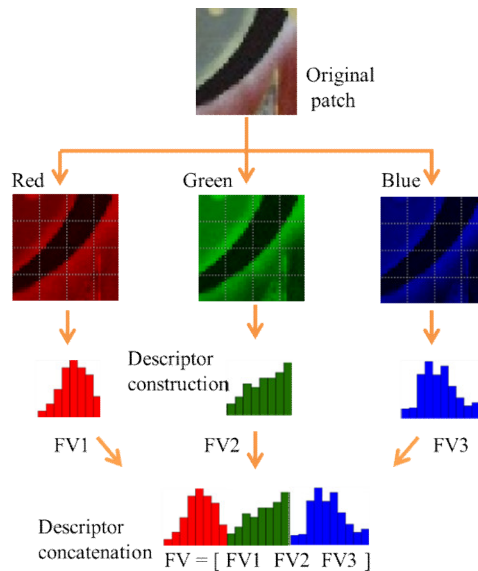

Figure. 8. Block diagram of RGB color descriptor construction, where FV1, FV2 and FV3 are descriptors built from red, green and blue regions respectively.

5   Experiments and Results

This section presents the performance evaluation of the proposed descriptors in four different applications: 1) texture classification, 2) image matching, 3) object recognition and 4) scene classification. First, the performance of the binary operators in texture classification is evaluated on

the Outex texture database[1]. It is a benchmark database highly used to evaluate the performance of many binary operators [15, 17, 18, 37, 58, 60]. The state-of-the-art binary operators LBP, LBPriu2, RLBP, CSLBP and OCLBP are used for comparative evaluation in texture classification. Second, the image matching performance of the proposed local image region descriptors is evaluated on challenging and widely used Oxford image matching dataset. The SIFT, LBP, CSLBP and OCLBP are compared with the proposed descriptors. SIFT is a widely used image region descriptor for image matching and object recognition. The comparative binary pattern based descriptors such as LBP, CSLBP and OCLBP gained much attention in image matching and object recognition recently because of their simplicity and effectiveness as mentioned in section 3. Third, the performance of proposed interest region descriptors in object recognition is evaluated on two widely used datasets such as SIMPLIcity dataset [61] and butterflies dataset [62]. Finally, the performance in scene classification is evaluated using the OT-scene dataset [63] that contains challenging image sets.

5.1   Texture classification performance on Outex database

The Outex datasets such as Outex-TC-00010 (TC10), Outex-TC-00011n (TC11n), Outex-TC-00012 (TC12) are used to evaluate texture classification performance of the proposed descriptors. The dataset TC10 contains 4320 images with illumination and rotation variations. There are 24 categories of images in TC10 among them first 20 images from each category are chosen for training where remaining images are used for testing. The dataset TC11n contains 960 images with Gaussian noise (σ=5), among these the first 20 images are considered as training images and the remaining images are taken as testing images. The TC12 contains 1440 images with illumination variations. From that, the first 480 images are used as training set and 960 images are considered as testing set. Sample images from the Outex database TC10, TC11n and TC12 are shown in Fig 9.
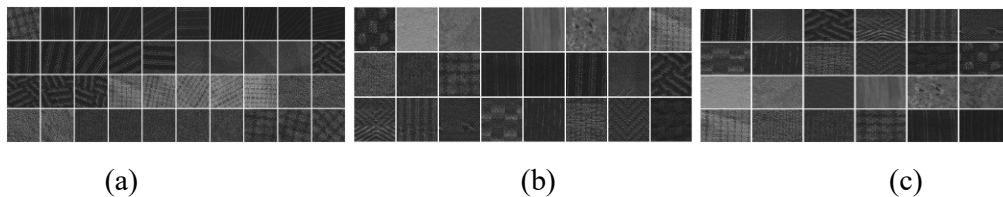

(a)   (b)   (c)

Figure. 9. Sample images from Outex texture dataset, where (a) Outex-TC-00010, (b) Outex-TC-00011n and (c) Outex-TC-00012

A nearest neighbor classifier is exploited for texture classification where the dissimilarity between two texture features are computed using chi-square distance measure as in [18]. The chi-square distance between two texture features is computed as follows. Assume that the $S = \{s_i\}$ and $M = \{m_i\}$ are any two binary pattern histograms, then the dissimilarity measure is calculated as:

$$\chi^2(S,M) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i} \tag{8}$$

The texture classification results for three parameter settings such as (R=1, N=8), (R=2, N=8) and (R=2, N=16) are given in the table 1. The binary pattern operators such as the LBP [17], LBPriu2 [37], CSLBP [15], OCLBP [18] and RLBP [58] are compared with the proposed texture operators LIOCSLBP and LIOOCLBP. Since the datasets consist of gray scale texture patches, the color texture descriptors of the proposed descriptors are not evaluated here. Table 1 depicts the classification accuracies of different texture operators evaluated under different parameter setups for the texture datasets.

---

[1]http://lagis-vi.univ-lille1.fr/datasets/outex.html

The average texture classification accuracy depicted in table 1 shows that the proposed texture operator LIOCSLBP outperforms other comparative state-of-the-art texture operators. The proposed LIOOCLBP also shows very competitive performance compared to other texture operators. Unlike the peer texture operators, the proposed descriptors LIOCSLBP and LIOOCLBP give consistent performance in all datasets under all the parameter setups. The results also show that the proposed LIOCSLBP performs better than the proposed LIOOCLBP.

For the TC10 dataset, the proposed LIOCSLBP descriptor outperforms other approaches for the parameter setup (R1, N8) where the accuracy is only slightly surpassed by LBPriu2 in other two parameter setups (R2, N8) and (R2, N16). For the TC11n dataset for noise, the proposed LIOCSLBP texture operator outperforms comparative approaches in parameter setups (R1, N8) and (R2, N16) where OCLBP outperformed LIOCSLBP only with a slightly higher accuracy. The proposed LIOCSLBP and LIOOCLBP texture operators present consistent performance across different parameter settings where the accuracies of the competitive approaches such as OCLBP, LBPriu2 and CSLBP drop as the parameters *R* and *N* increases correspondingly. Among the comparative approaches, the RLBP performs better than the LBP, CSLBP and OCLBP for noisy images since it uses the average local gray level value instead of center pixel value to compute binary patterns. For TC12 dataset for illumination, the original LBP outperforms other approaches in almost all parameter settings. However the proposed texture operator LIOCSLBP achieves appreciable classification accuracies for the given parameter setups where the accuracy is very closer to the comparative approaches in the (R2, N16) parameter setup.

Methods	R1, N8			R2, N8			R2, N16			Average
LBP [17]	50.05	18.75	**98.65**	52.03	20.42	**99.79**	53.03	42.08	99.79	59.40
LBPriu2 [37]	83.52	08.54	72.19	**84.08**	20.62	80.62	**88.85**	14.37	82.81	59.51
CSLBP [15]	51.87	27.50	78.64	52.13	34.58	79.06	52.68	44.79	83.33	56.06
OCLBP [18]	50.31	13.54	97.50	51.12	**51.46**	97.71	53.59	45.62	87.39	60.92
RLBP [58]	52.24	19.17	88.85	53.41	49.58	99.48	66.93	49.58	**99.89**	64.35
LIOCSLBP	**84.89**	**31.25**	78.64	82.84	51.25	83.33	85.73	**60.21**	99.79	**73.10**
LIOOCLBP	72.63	11.67	70.42	66.93	23.33	85.83	74.92	53.33	85.83	60.54

Table 1. Texture Classification performance (%) on Outex TC10, TC11n and TC12 (R – Radius, N – Number of neighbors)

The number of neighboring pixels (*N*) and radius (*R*) are the controlling parameters that decide the discriminative power, feature dimension and computation complexity of the proposed binary operators. The influence of these parameters can be directly assessed from the table 1. For the proposed descriptors, the binary pattern computation by setting *R* = 1 is not that effective since neighborhood pixels and its center pixel do not have much texture variations. The binary patterns computed with *R* = 2 and *N* = 16 are highly discriminative but computationally intensive and consume more memory. It is obvious that the texture classification accuracy improves better when the number of neighborhood pixels increases. Also, the feature dimension increases when the number of neighboring pixels increases, which leads to more computation complexity. Hence, the parameter setup *R* = 2 with *N* = 8 is suggested for balancing between the speed and memory to achieve moderate discriminative power. This parameter setting is further used in image matching, object recognition and scene recognition experiments.

5.2 Experiments on Image Matching

This section presents the details about the dataset, evaluation criteria and experimental results for image matching.

5.2.1 Dataset and evaluation criteria

The standard Oxford image matching dataset[2] is utilized to evaluate the performance of the proposed interest region descriptors. The dataset consist of images with different geometric and photometric variations including illumination, blur, viewpoint, scale and rotation. Sample images from the Oxford image matching dataset are shown in Fig. 10. The image matching evaluation criteria presented in [6] is followed to evaluate the performance of the interest region descriptors. The numbers of correct and false matches are observed for each image pair to evaluate the feature matching performance. The keypoint features of first image in each image category are matched against the keypoint features of second image in the same category as done in [40]. The Euclidean distance is used to find the matching regions with a distance threshold. Overlap error t is set to find the number of correct matches. Typically, the value of t is set to below 50% to find matching features. Then, the number of correct matches and false matches are counted. The results of the descriptors are presented in graph with recall versus 1-precision curve. The recall and 1-precision are calculated using the following equations.

$$recall = \frac{\#correct\ matches}{\#correspondences} \tag{9}$$

$$1 - precision = \frac{\#\ false\ matches}{\#correct\ matches + \#\ false\ matches} \tag{10}$$

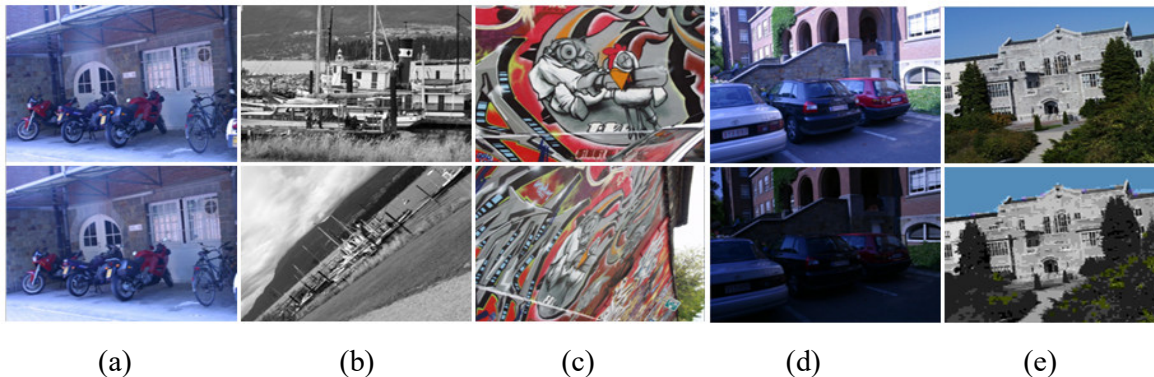(a)   (b)   (c)   (d)   (e)

Figure. 10. Sample images from the Oxford image matching dataset. (a) Bike (blur), (b) Boat (scale and rotation), (c) Graf (viewpoint), (d) Leuven (illumination) and (e) UBC (compression).

5.2.2 Performance on image matching

The proposed descriptors are compared with the state-of-the-art descriptors such as SIFT [1], LBP [17], CSLBP [15] and OCLBP [18]. By following [15, 18, 39] parameters R and N are set to 2, 8 respectively for the LBP, CSLBP, OCLBP, LIOCSLBP, LIOOCLBP, RGBLIOCSLBP, HSVLIOCSLBP, RGBLIOOCLBP and HSVLIOOCLBP descriptors. Therefore, the dimensionalities of the above said descriptors using single interest region around for each interest point are 4096, 256, 512, 256, 512, 768 and 1536 respectively.
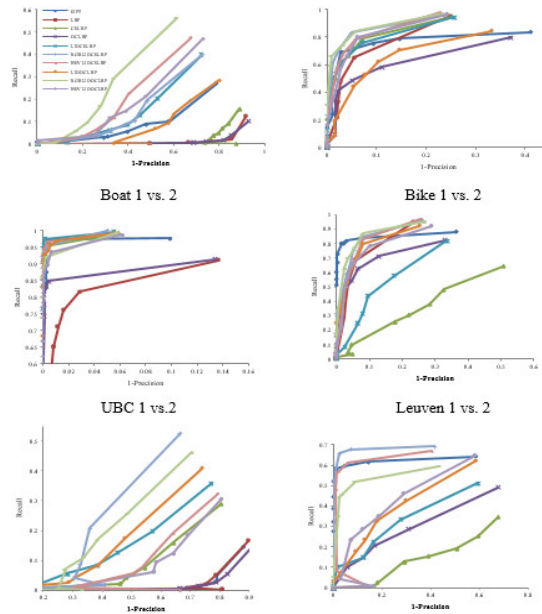
Figure. 11. Image matching performance on the Oxford dataset

For evaluation, first and second images of the image sequence from the Oxford dataset are considered similar to [18, 40]. The performance of the proposed descriptors is tested for images with geometric (rotation, scale, and viewpoint) and photometric transformations (blur and JPEG compression). The results of the proposed descriptors compared with the state-of-the-art descriptors are shown in Fig. 11. The results show that the proposed color enhanced descriptors RGBLIOCSLBP, HSVLIOCSLBP, RGBLIOOCLBP and HSVLIOOCLBP show the best performances than the state-of-the-art descriptors such as SIFT, LBP, CSLBP and OCLBP for images with photometric and geometric challenges. Also the proposed descriptors LIOCSLBP and LIOOCLBP show competitive performances to the other comparative descriptors. It can be concluded that the incorporation of color information in the descriptor increases the illumination invariance. Although the incorporation of color information increased dimensionality that improves the discriminative power of the descriptor which helps one-to-one keypoint matching.

5.3   Experiments on object recognition

The keypoint descriptors are used for object recognition by exploiting the BOVW approach. The SIMPLIcity [61] and Butterflies [62] datasets are used to evaluate the performance of the proposed descriptors in object recognition. The construction of the BOVW for image representation is presented in Fig. 12.
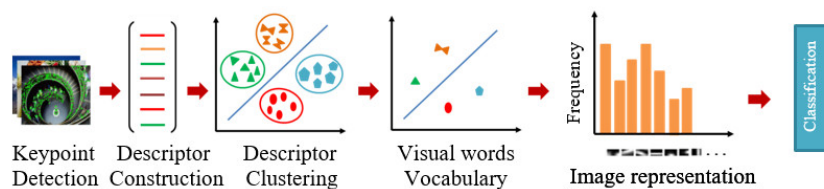

Figure. 12. Bag-of-visual-words construction and image representation

5.3.1 Feature extraction

Initially, the Harris keypoint detector [13] is used to find the interest points from images. Then features (SIFT, LBP, CSLBP, OCLBP, LIOCSLBP, RGBLIOCSLBP, HSVLIOCSLBP, LIOOCLBP, RGBLIOOCLBP and HSVLIOOCLBP) are extracted from the local region around each interest point. Then, the bag-of-visual words approach is applied for image classification.

5.3.2 Bag-of-Visual-Words (BOVW) modeling

After the features are extracted, the BOVW approach is applied to represent each image in terms of a histogram of visual dictionary. The number of local features extracted from one image differs to another image, where the BOVW model represents the image as a histogram of keypoints with standard dimension. The BOVW constructs an order-less collection of local image patch descriptions called *visual vocabulary* (or dictionary) using clustering methods. Here, the K-means clustering with squared Euclidean distance measure is exploited to cluster the local features where the set of cluster centers are considered as visual-word vocabulary. These visual vocabularies are generated for all descriptors separately.

Setting the size of the visual-word vocabulary is an important task, since small size leads to less discriminative features where very large size leads to less generalizable features and is time consuming. In order to maintain the trade-off between discrimination and generalization, the visual vocabulary size is chosen accordingly. Using the constructed visual words, feature vectors of an image keypoints are then quantized into their closest visual word. A histogram is computed for each image according to the number of descriptions assigned using the corresponding visual words. The resultant features are compact, informative and of fixed–length representation [18], characteristics which are further used for image classification.

5.3.3 Classification

The images represented by the BOVW features are further used for classification. Here, the Support Vector Machine (SVM) [64] algorithm is utilized for image classification.

5.3.4 Performance on butterflies dataset

The Butterflies dataset [62] contains 619 images with 7 categories of butterflies such as Admiral, Black Swallow tail, Machaon, Monarch open, Monarch closed, Peacock and Zebra, where each category contains 111, 42, 83, 84, 74, 134 and 91 images respectively. Sample images from the dataset are shown in Fig. 13. Each butterfly category contains images with motion blur, affine, rotation, illumination and translation changes. First 10 images from each butterfly category are chosen to generate a visual vocabulary. A visual vocabulary containing 500 visual words is generated using K-means clustering algorithm for each descriptor.

The classification results for each butterfly category are presented in table 2. The proposed LIOCSLBP and LIOOCLBP descriptors present better performance than the start-of-the-art region descriptors. The proposed color descriptor HSVLIOCSLBP outperforms other descriptors in most of the cases. Color is an important clue for recognizing butterflies as each type of butterfly has unique color patterns. That is the reason behind the high classification accuracy of the proposed color descriptors. It is discovered that the LBP, CSLBP and OCLBP descriptors lose their discriminative power due to large illumination and rotation variations in the images. It should be also noted that the proposed gray scale descriptors LIOCSLBP and LIOOCLBP outperform comparative approaches for many butterfly categories.

Figure. 13. Sample images from the Butterflies dataset. Top row contains categories of admiral, Machaon and Monarch_closed (two images per category). Bottom two rows contain Monarch open, Peacock, Black_swallowtail and Zebra (three images per category)

Methods	Admiral	Black swallow tail	Machaon	Monarch open	Monarch closed	Peacock	Zebra	Average
SIFT [1]	51.79	33.33	52.38	56.76	54.76	61.19	77.78	55.43
LBP [17]	58.93	28.57	47.62	45.95	42.86	71.64	66.67	51.75
CSLBP [15]	60.71	19.05	59.52	27.02	64.29	53.73	55.56	48.55
OCLBP [18]	37.5	52.38	50.00	32.43	59.52	53.73	64.44	50.00
LIOCSLBP	62.5	52.38	64.29	51.35	59.52	67.16	66.67	60.55
RGBLIOCSLBP	64.29	47.62	61.91	64.86	64.29	64.18	77.78	63.56
HSVLIOCSLBP	**89.29**	**71.43**	54.76	**86.49**	**69.05**	**97.01**	68.69	**76.67**
LIOOCLBP	60.71	47.62	**80.95**	45.95	64.29	91.05	75.56	66.59
RGBLIOOCLBP	87.5	38.09	71.43	67.57	52.38	88.06	66.67	67.39
HSVLIOOCLBP	73.21	63.27	63.27	63.27	63.27	63.27	63.27	63.27

Table 2. Classification performance on Butterfly dataset

5.3.5 Object recognition on SIMPLIcity dataset

The SIMPLIcity dataset [61] is the subset of COREL image database. This subset contains 1000 images with 10 different categories named African people, beach, building, bus, elephant, flower, food, horse, dinosaur, and mountain. The size of the images is either 256×384 or 384×256 pixels. Sample images from the SIMPLIcity dataset are shown in Fig. 14. Following [18], a vocabulary of 1000 visual words is generated using K-means for each descriptor. Using these visual words, all the dataset images are represented with equal sized histogram. From each category, first 50% of the images are taken for training and remaining images are taken as testing images.

Figure. 14. Sample images from the SIMPLICity dataset for categories such as African People, Beach, Building, Bus, Dinosaur, Elephant, Flower, Horse and Mountain and Food (from top left to bottom right).

The object categorization accuracies of proposed descriptors and comparative descriptors are presented in table 3. The proposed descriptor LIOOCLBP presents the best results for the SIMPLIcity dataset in terms of average accuracy. Both LIOCSLBP and LIOOCLBP shows better and competitive performance than the SIFT, LBP, CSLBP and OCLBP descriptors for most image categories. The color enhanced descriptors such as RGBLIOCSLBP, HSVLIOCSLBP, RGBLIOOCLBP and RGBLIOOCLBP also performed better than gray image based descriptors in some cases.

Descriptors	People	Beach	Building	Bus	Elephant	Flower	Food	Horse	Dinosaur	Mountain	Average
Table. 3. Object recognition performance on SIMPLIcity dataset											
SIFT [1]	66	52	44	90	**100**	**78**	74	**88**	46	38	67.6
LBP [17]	40	48	66	96	90	26	52	66	42	58	58.4
CSLBP [15]	66	36	50	76	**100**	36	56	70	42	48	58.0
OCLBP [18]	64	**54**	64	**98**	96	42	56	68	46	66	65.4
LIOCSLBP	72	**54**	66	**98**	96	52	64	70	56	**74**	70.2
RGBLIOCSLBP	**86**	42	**68**	88	98	56	72	82	40	62	69.4
HSVLIOCSLBP	66	42	48	82	**100**	56	50	70	48	60	62.2
LIOOCLBP	70	52	**68**	94	96	54	**78**	68	**64**	70	**71.4**
RGBLIOOCLBP	68	46	40	94	98	66	62	84	50	70	67.8
HSVLIOOCLBP	64	52	56	92	98	62	68	76	58	64	69.0

5.3.6 Experiments on Scene recognition

The OT-scene dataset [63] is used to evaluate the performance of the descriptors in scene recognition. The dataset consists of eight scene categories such as coast, forest, highway, inside city, mountain, open country, street and tall building where each class contains 360, 328, 260, 308, 374, 410, 292 and 356 images respectively. The size of the images in the dataset is 256×256 pixels. Fig.15 depicts sample images of different categories from the dataset. The dataset contains images with large variations in illumination which makes it a challenging dataset for descriptor evaluation. The BOVW modeling is used with k = 2000 based on the work in [18]. From each category, 50% of images are considered as the training image set and the remaining are treated as the test image set.

Figure. 15. Sample images from the OT scene dataset. Top row images - Coast, Forest, Highway, Inside city. Bottom row images - Mountain, Open country, Street and Tall building.
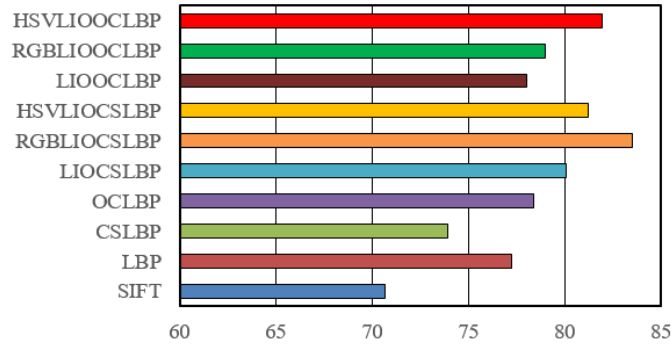

Figure. 16. Scene classification results on OT scene dataset

The observed results for scene classification are presented in Fig. 16. The proposed color descriptor LIOCSLBP presents best performance in scene classification for OT scene dataset. It can be seen that the proposed LIOCSLBP and LIOOCLBP descriptors achieve better and very competitive performance than the CSLBP and OCLBP descriptors. Also, they pose better recognition accuracy than the SIFT and LBP descriptors. Since color is an important cue in recognizing differences scenes, the color descriptors achieve better performance than the gray value based descriptors which lose color information. This also leads the proposed gray image based descriptors LIOCSLBP and LIOOCLBP to pose comparatively lower recognition accuracy when there are large illumination changes in images where the color enhanced descriptors express the strongest invariant property to illumination changes. Thus, the proposed color enhanced descriptors display better performance than the descriptors that used gray image for description.

Table 4. Computation time comparison (in seconds)			
Descriptors	Oxford dataset	SIMPLIcity dataset	OT Scene dataset
SIFT [1]	2.1470	0.5102	0.3112
LBP [17]	0.3103	0.1032	0.1064
CSLBP [15]	0.1851	0.0512	0.0492
OCLBP [18]	0.3812	0.1143	0.0824
LIOCSLBP	0.5577	0.1254	0.0852
RGBLIOCSLBP	0.7332	0.3124	0.1126
HSVLIOCSLBP	0.7410	0.3296	0.1032
LIOOCLBP	0.7562	0.1882	0.1444
RGBLIOOCLBP	1.2016	0.4874	0.2984
HSVLIOOCLBP	1.2648	0.4794	0.3088

5.4 Computation cost

This section presents the computation time of the proposed descriptors and the state-of-the-art descriptors which are used for comparison in the previous sections. A personal computer running the windows 7 operating system with a Pentium Dual-Core CPU T4400 @ 2.20GHz + 2.20GHz and 3GB RAM is used to compute the computation time. The MATLAB 2014a tool (version 8.3.0.532) is used to implement the proposed interest region descriptors. The average computation time taken per image for each descriptor is presented in Table 4. From Table 4, it could be understood that the proposed descriptors are computationally inexpensive compared to the CSLBP and OCLBP descriptors respectively due to the additional step LIO. However, they are less time consuming than the SIFT descriptor. Due to the higher dimensionality of the color descriptors, they are slower than their corresponding gray scale descriptors. The proposed color descriptors take comparatively more time than the proposed gray scale descriptors due to their higher dimensionality. Moreover, the computation times of the proposed approaches are still appreciable and can be used in many applications where the speed can be optimized using code optimization.

6   Conclusion

In this paper, two novel binary pattern based rotation invariant interest region descriptors denoted as LIOCSLBP and LIOOCLBP are proposed. The proposed descriptors aims at computing rotation invariant binary patterns for region description using the Local Intensity Ordering rather than exploiting dominant orientation estimation. In contrast to most of the previous descriptors, the proposed descriptors used double interest regions centering on each interest point to build features. Hence, the features built by the proposed descriptors are highly discriminative and avoid ambiguity in keypoint matching. In order to further improve the discriminative ability, four color enhanced descriptors such as RGBLIOCSLBP, RGBLIOOCLBP, HSVLIOOCLBP and HSVLIOCSLBP are also proposed using RGB or HSV color models. These color descriptors further increase discriminative power, improve invariance to illumination changes.

An extensive evaluation comparing proposed descriptors with state-of-the-art descriptors was conducted on popular image datasets for applications such as texture classification, image matching, object recognition and scene recognition. The experimental results have revealed the efficiency of the proposed descriptors in delivering better results than the state-of-the-art descriptors such as SIFT, LBP, CSLBP and OCLBP. The experimental results also proved that the simple operation of Local Intensity Ordering can achieve reliable and efficient rotation invariant binary pattern computation. The results have shown that incorporation of color information in keypoint description improves the feature discrimination ability where the proposed color descriptors outperformed both comparative and proposed gray descriptors in most cases. It is recommended that the choice between using gray scale and color descriptors can be determined based on the speed and accuracy requirements of a specific application. In future, the proposed descriptors will be applied to interesting applications such as large scale image retrieval. Moreover, the proposed local image region descriptor approach will also be extended to use the local ternary patterns.

REFERENCES

[1]   Lowe D. G (2004) Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2):91-110.

[2]   Zhang J, Marszałek M, Lazebnik S, Schmid C (2007) Local features and kernels for classification of texture and object categories: A comprehensive study. International journal of computer vision 73(2):213-238.

[3]   Szeliski R (2006) Image alignment and stitching: A tutorial. Foundations and Trends in Computer Graphics and Vision 2(1):1-104.

[4]   Fan B, Wu F, Hu Z (2012) Rotationally invariant descriptors using intensity order pooling. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(10):2031-2045.

[5]   Ke Y, Sukthankar R (2004) PCA-SIFT: A more distinctive representation for local image descriptors. In: Proceedings of Computer Vision and Pattern Recognition, pp 506-513.

[6]   Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence 27(10):1615-1630.

[7]   Bay H, Tuytelaars T, Van Gool L (2006) Surf: Speeded up robust features. In: Proceedings of European conference on computer vision, pp 404-417.

[8]   Tola E, Lepetit V, Fua P (2010) Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE transactions on pattern analysis and machine intelligence 32(5):815-830.

[9]   Calonder M, Lepetit V, Ozuysal M, Trzcinski T, Strecha C, Fua P (2012) BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7):1281-1298.

[10]   Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: An efficient alternative to SIFT or SURF. In: 2011 International conference on computer vision, pp 564-2571.

[11]   Wang Z, Fan B, Wu F (2014). Affine subspace representation for feature description. In: Proceedings of European Conference on Computer Vision, pp 94-108.

[12]   Lindeberg T (1998) Feature detection with automatic scale selection. International journal of computer vision 30(2):79-116.

[13]   Mikolajczyk K, Schmid C (2004) Scale & affine invariant interest point detectors. International journal of computer vision 60(1):63-86.

[14]   Mikolajczyk K, et al., (2005) A comparison of affine region detectors. International journal of computer vision 65(1-2):43-72.

[15]   Heikkila M, Pietikäinen M, Schmid C (2009) Description of interest regions with local binary patterns. Pattern recognition 42(3):425-436.

[16]   Wang Z, Fan B, Wu F (2011) Local intensity order pattern for feature description. In: Proceedings of International Conference on Computer Vision, pp 603-610.

[17]   Ojala T, Pietikainen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern recognition 29(1):51-59.

[18]   Zhu C, Bichot C. E, Chen L (2013) Image region description using orthogonal combination of local binary patterns enhanced with color information. Pattern Recognition 46(7):1949-1963.

[19]   Qi X, Lu Y, Chen S, Li C. G, Guo J (2013) Spatial co-occurrence of local intensity order for face recognition. In: Proceedings of Multimedia and Expo Workshops, pp 1-6.

[20]   Kang T. K, Choi I H, Lim M. T (2015) MDGHM-SURF: A robust local image descriptor based on modified discrete Gaussian–Hermite moment. Pattern Recognition 48(3):670-684.

[21]   Berg A. C, Malik J (2001) Geometric blur for template matching. In: Proceedings of Computer Vision and Pattern Recognition, 1: I-607.

[22]   Chen J, Shan S, Zhao G, Chen X, Gao W, Pietikainen M (2008) A robust descriptor based on weber's law. In: Proceedings of Computer Vision and Pattern Recognition, pp 1-7.

[23]   Kannan R, Ghinea G, Kannaiyan S, Swaminathan S (2014) MPRF: Multisupport polar region features for image description. In: Proceedings of International Symposium on Signal Processing and Information Technology, pp 000049-000054.

[24]   Winder S. A, Brown M (2007) Learning local image descriptors. In: Proceedings of Conference on Computer Vision and Pattern Recognition, pp 1-8.

[25]   Mittal A, Ramesh V (2006) An intensity-augmented ordinal measure for visual correspondence. In: Proceedings of Computer Vision and Pattern Recognition, 1:849-856.

[26]   Gupta R, Patil H, Mittal A (2010) Robust order-based methods for feature description. In: Proceedings of Computer Vision and Pattern Recognition, pp 334-341.

[27]   Tang F, Lim S. H, Chang N. L, Tao, H (2009) A novel feature descriptor invariant to complex brightness changes. In: Proceedings of Computer Vision and Pattern Recognition, pp 2631-2638.

[28]   Kannan R, Ghinea G, Kannaiyan S, Swaminathan S (2015) CLRF: Compressed Local Retinal Features for Image Description. In: Proceedings of Advances in Pattern Recognition, pp 1-5.

[29]   Dubey S. R, Singh S. K, Singh R. K (2014) Rotation and illumination invariant interleaved intensity order-based local descriptor, IEEE Transactions on Image Processing, 23(12):5323-5333.

[30]   Yan P, Liang D, Tang J, Zhu M (2016) Local feature descriptor using entropy rate. Neurocomputing, 194:157-167.

[31]   Yan P, Liang D, Tang J, Zhu M (2016) Local feature descriptor invariant to monotonic illumination changes. Journal of Electronic Imaging, 25(1):013023-013023.

[32]   Huang D, Zhu C, Wang Y, Chen L (2014). HSOG: a novel local image descriptor based on histograms of the second-order gradients. IEEE Transactions on Image Processing, 23(11), 4680-4695.

[33]   Tian T, Sethi I, Ming D, Patel N (2015) A Zoned Image Patch Permutation Descriptor. IEEE Signal Processing Letters, 22(6): 728-732.

[34]   Bosch A, Zisserman A, Munoz X (2008) Scene classification using a hybrid generative/discriminative approach. IEEE transactions on pattern analysis and machine intelligence, 30(4):712-727.

[35]   Verma M, Raman B (2015) Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval. Journal of Visual Communication and Image Representation, 32:224-236.

[36]   Dubey S. R, Singh S. K, Singh R. K (2015) Local diagonal extrema pattern: a new and efficient feature descriptor for CT image retrieval. IEEE Signal Processing Letters, 22(9):1215-1219.

[37]   Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7):971-987.

[38]   Liao S, Law M. W, Chung A. C (2009) Dominant local binary patterns for texture classification. IEEE transactions on image processing, 18(5):1107-1118.

[39]   Zhao Y, Huang D. S, Jia W (2012) Completed local binary count for rotation invariant texture classification. IEEE transactions on image processing, 21(10):4492-4497.

[40]   Huang M, Mu Z, Zeng H, Huang S (2015) Local image region description using orthogonal symmetric local ternary pattern. Pattern Recognition Letters, 54:56-62.

[41]   Qi X, Xiao R, Li C. G, Qiao Y, Guo J, Tang X (2014) Pairwise rotation invariant co-occurrence local binary pattern. IEEE transactions on pattern analysis and machine intelligence, 36(11):2199-2213.

[42]   Yang X, Cheng K. T. T. (2014) Local difference binary for ultrafast and distinctive feature description. IEEE transactions on pattern analysis and machine intelligence, 36(1):188-194.

[43]   Shang J, Chen C, Pei X, Liang H, Tang H, Sarem M, (2015) A novel local derivative quantized binary pattern for object recognition. The Visual Computer, 1-13.

[44]   Li Y, Tan J, Zhong J, Chen Q (2016) Compact descriptor for local feature using dominating centre-symmetric local binary pattern. IET Computer Vision, 10(1):36-42.

[45]   Abdel-Hakim, A. E, Farag A. A (2006) CSIFT: A SIFT descriptor with color invariant characteristics. In: Proceedings of Computer Vision and Pattern Recognition, 2:1978-1983.

[46]   Van de Weijer J, Gevers T, Bagdanov A. D (2006) Boosting color saliency in image feature detection. IEEE transactions on pattern analysis and machine intelligence, 28(1):150-156.

[47]   Burghouts G. J, Geusebroek J. M (2009) Performance evaluation of local colour invariants. Computer Vision and Image Understanding, 113(1):48-62.

[48]   Van De Sande K, Gevers T, Snoek C (2010) Evaluating color descriptors for object and scene recognition. IEEE transactions on pattern analysis and machine intelligence, 32(9):1582-1596.

[49]   Bach F. R, Lanckriet G. R, Jordan M. I (2004) Multiple kernel learning, conic duality, and the SMO algorithm. In: Proceedings of the twenty-first international conference on Machine learning, (p. 6).

[50]   Sivic J, Zisserman A (2003) Video Google: A text retrieval approach to object matching in videos. In: Proceedings of International Conference on Computer Vision, pp 1470-1477.

[51]   Kavukcuoglu K, Sermanet P, Boureau Y. L, Gregor K, Mathieu M, Cun Y. L (2010) Learning convolutional feature hierarchies for visual recognition. In: Proceedings of Advances in neural information processing systems, pp 1090-1098.

[52]   Goh H, Thome N, Cord M, Lim J. H (2012) Unsupervised and supervised visual codes with restricted boltzmann machines. In: Proceedings of European Conference on Computer Vision, pp 298-311. Springer Berlin Heidelberg.

[53]   Vedaldi A, Gulshan V, Varma M, Zisserman A (2009) Multiple kernels for object detection. In: Proceedings of international conference on computer vision, pp 606-613.

[54]   Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of Computer Vision and Pattern Recognition, 2:2169-2178.

[55]   Perronnin F, Dance C (2007) Fisher kernels on visual vocabularies for image categorization. In: Proceedings of Computer Vision and Pattern Recognition, pp 1-8.

[56]   Arandjelovic R, Zisserman A (2013) All about VLAD. In: Proceedings of Computer Vision and Pattern Recognition, pp 1578-1585.

[57]   Zhou X, Yu K, Zhang T, Huang T. S (2010) Image classification using super-vector coding of local image descriptors. In: Proceedings of European conference on computer vision, pp 141-154. Springer Berlin Heidelberg.

[58]   Zhao Y, Jia W, Hu R. X, Min H (2013) Completed robust local binary pattern for texture classification. Neurocomputing 106:68-76.

[59]   Lazebnik S, Schmid C, Ponce J (2005) A sparse texture representation using local affine regions. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8):1265-1278.

[60]   He Y, Sang N, Gao C (2013) Multi-structure local binary patterns for texture classification. Pattern Analysis and Applications 16(4):595-607.

[61]   Wang J. Z, Li J, Wiederhold G (2001) SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on pattern analysis and machine intelligence 23(9):947-963.

[62]   Lazebnik S, Schmid C, Ponce J (2004) Semi-local affine parts for object recognition. In: Proceedings of British Machine Vision Conference, pp 779-788.

[63]   Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42(3):145-175.

[64]   Cortes C, Vapnik, V (1995) Support-vector networks. Machine learning, 20(3):273-297.