Color orthogonal local binary patterns combination for image region description

(Technical Report)

Chao Zhu, Charles-Edmond Bichot, Liming Chen
Université de Lyon, CNRS
Ecole Centrale de Lyon, LIRIS, UMR5205, F-69134, France
E-mail: {chao.zhu, charles-edmond.bichot, liming.chen}@ec-lyon.fr

Abstract—**Visual content description is a key issue for machine-based image analysis and understanding. A good visual descriptor should be both discriminative enough and computationally efficient while possessing some properties of robustness to viewpoint changes and lighting condition variations. In this paper, we propose several new local descriptors based on color orthogonal local binary patterns combination (OLBPC) for image region description. The aim is to increase both discriminative power and photometric invariance properties of the original LBP while keeping its computational efficiency. The experiments in three different applications show that the proposed descriptors outperform the popular SIFT and CS-LBP, and get comparable or even slightly better performances than the state-of-the-art color SIFT descriptors. Meanwhile, they could provide complementary information to the color SIFT, because a fusion of these two kinds of descriptors is found to perform clearly better than either of the two separately. Moreover, the proposed descriptors are more computationally efficient than the color SIFT (about 2 times faster).**

Keywords- Local descriptor; Region description; Orthogonal local binary patterns combination; Color LBP descriptor; CS-LBP; SIFT; Image matching; Object recognition; Scene classification

1. Introduction

One of the most challenging problems in computer vision is content-based high-level image analysis and understanding, such as object recognition or scene classification, mainly due to intra-class variations and inter-class similarities. Therefore, a key issue for such tasks is to generate effective visual content descriptions, which could ideally be discriminative enough, computationally efficient, and possessing some properties of robustness to changes in viewpoint, scale and lighting conditions.

Earlier work in this domain has mainly utilized global features as image descriptions, including color histogram [1], color moments [2], edge histogram [3], texture co-occurrence matrix [4], and so on. These features are extracted directly from the whole image, thus encoding the global visual content of an image. While quite efficient to compute, the downside of these global features is their great sensitivity to clutter, occlusion, viewpoint changes and illumination variations.

For this reason, global features have gradually given way later on to local image descriptors, which have received a lot of attention in recent years, and have already gained the popularity and dominance in image content analysis related tasks nowadays. Instead of from the whole image, the idea of local image descriptors is to extract features from local image regions centered either on some sparse keypoints with certain invariance properties, for instance with respect to scale and viewpoint change, or simply on a dense sampling grid. By this way, local image descriptors could be more powerful in discrimination and be more robust to image variations, compared with the global ones.

Many different local image descriptors have been proposed in the literature, and the most popular ones are distribution-based descriptors, represented by the SIFT [5], which is a 3D histogram of gradient locations and orientations. The location is quantized into a 4 by 4 location grid and the gradient angle is quantized into 8 orientations, resulting in a 128-dimensional descriptor. The contributions to the gradient orientations are weighted by the gradient magnitudes and a Gaussian window overlaid over the region, indicating that more emphases are put on the gradients near the region center. Other widely used distribution-based local descriptors include PCA-SIFT [6], GLOH [7], SURF [8] and HOG [9]. They are all related to the SIFT, and can be considered as the extension or refinement of the original SIFT.

Several comprehensive studies on local image descriptors [7,10,11] have shown that distribution-based descriptors perform significantly better than other features, and achieve the best results in tasks as diverse as image region matching, texture classification, object recognition and scene classification. Among them, the SIFT has been proven to be the most powerful and successful, and has been widely applied as the dominant feature in the state-of-the-art recognition/classification systems [12]. Moreover, since the SIFT is an intensity based descriptor without color information, several color SIFT descriptors have been proposed [13,14,15,16] to increase its discriminative power. In [17], the authors evaluated different color descriptors in a structured way, and recommended to use color SIFT descriptors for object and scene recognition because they outperform the SIFT. However, the downside of the color SIFT descriptors is their high computational cost, especially when the size of image or the scale of dataset increases significantly. Therefore, a new local descriptor would be preferred if it could be more computationally efficient, and keep high discriminative power at the same time.

The local binary pattern (LBP) operator [18] is well known as a good texture feature, and has been successfully applied for many applications, such as texture classification [19,20,21], texture segmentation [22], face recognition [23], and facial expression recognition [24]. Its advantage of computational simplicity and good power for texture structure description makes it a good candidate for describing local image regions. The bottleneck lies in the high dimensional feature vectors produced by the LBP, especially when more neighboring pixels are taken into consideration. The so-called "curse of dimensionality" will be caused if they are used directly to build a local region descriptor. In [25], the authors proposed center-symmetric local binary pattern (CS-LBP) to reduce the size of the original LBP histogram and use it for local region description. Another way to reduce the LBP dimension is to use the "uniform patterns" [19]. In this paper, we propose a new dimensionality reduction method for the LBP, denoted as orthogonal local binary patterns combination (OLBPC), which is more suitable for local descriptor construction. The basic idea is firstly splitting the neighboring pixels of the original LBP into several non-overlapped orthogonal groups, then computing the LBP separately for each group, and finally concatenating them together. The experimental results show that our method is more efficient than both CS-LBP and "uniform patterns" for the LBP dimensionality reduction, because it could keep high discriminative power with the smallest histogram size.

We then adopt the OLBPC operator in a SIFT-like approach to build a new descriptor, denoted as OLBPC descriptor, for local image regions. Generally speaking, the building process consists of the following steps: given local regions of an image, dividing each region into small cells for spatial information; in each cell, computing the OLBPC feature for each pixel and then building a histogram; concatenating all the histograms from each cell as the final descriptor for the region.

Derived from the LBP, the OLBPC descriptor is also an intensity-based descriptor without any color information of the image, while color plays an important role for distinction between objects, especially in natural scenes. Also, there can be various changes in lighting and viewing conditions in real-world scenes, leading to large variations of objects in surface illumination, and making recognition tasks more complicated and challenging. According to its definition, the OLBPC descriptor is only invariant to monotonic light changes in gray-level, and is deficient in power for dealing with these variations which mostly occur in natural scenes. Therefore, inspired by the color SIFT descriptors, we extend the OLBPC descriptor to different color space and propose several color OLBPC descriptors in this paper to increase its photometric invariance property and discriminative power. The experimental results in three different applications show that the color OLBPC descriptors outperform the popular SIFT and CS-LBP, and get comparable or even slightly better performances than the state-of-the-art color SIFT descriptors. Meanwhile, they could provide complementary information to the color SIFT, because a fusion of these two kinds of descriptors is found to perform clearly better than either of the two separately. Moreover, the proposed descriptors are more computationally efficient than the color SIFT (about 2 times faster).

The remaining sections are organized as follows. Section 2 introduces the proposed orthogonal local binary patterns combination (OLBPC) operator in detail, and compares it with other two popular LBP dimensionality reduction methods — "uniform patterns" and CS-LBP. The construction of the OLBPC descriptor for local image regions is described in section 3. We then give details of the proposed color OLBPC descriptors in section 4, including illumination change modeling and invariance property analysis for each descriptor. Section 5 presents the experimental evaluation of the proposed descriptors in three different applications. Finally, we conclude the paper in section 6.

2. Dimensionality reduction for the LBP

2.1. Original LBP operator

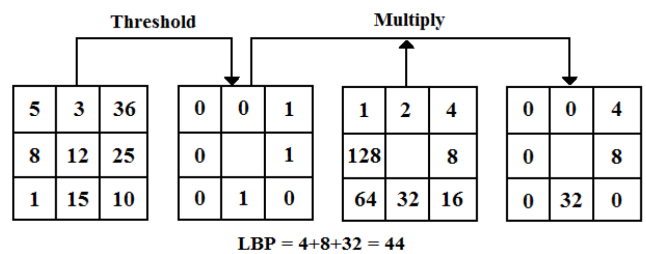

Fig. 1 Calculation of the original LBP operator

The original LBP operator can be seen as a unified approach to statistical and structural texture analysis. Fig. 1 gives an example. For one pixel in a gray image, its eight neighboring pixels are considered — their values are operated by the value of the central pixel as threshold. Precisely, for each neighboring pixel, the result will be set to one if its value is no less than the value of the central pixel, otherwise the result will be set to zero. The LBP code of the central pixel is then obtained by multiplying the results with weights given by powers of two, and summing them up together. Then, the LBP operator is extended to use a circular neighborhood with variant radius and variant number of neighboring pixels. Accordingly, the LBP code of the pixel at (x_c, y_c) is calculated by the following equation:

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} S(g_p - g_c) \times 2^p, S(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (1)$$

where g_p is the value of the neighboring pixel, g_c is the value of the central pixel, and P is the total number of the neighboring pixels.

For each pixel in an image, the same process is followed to get its LBP code, and the final LBP feature is obtained by building a histogram based on these codes. It can be seen that the LBP is very fast to calculate, and is invariant to monotonic illumination changes. Thus it is a good candidate for local image region description.

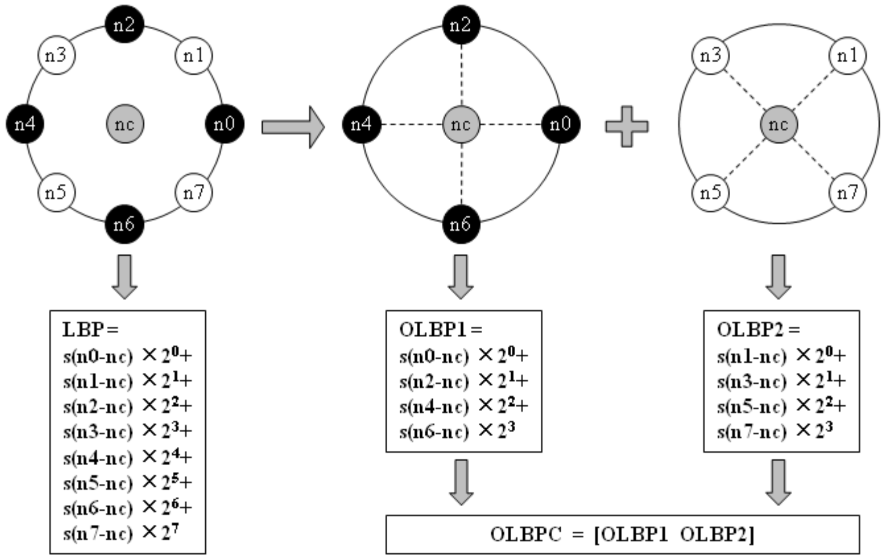

Fig. 2 Calculation of the LBP and OLBPC with 8 neighboring pixels

2.2. Orthogonal local binary patterns combination (OLBPC)

As introduced in the previous section, the LBP operator has its own advantages to be used for local region description. But the bottleneck lies in the high dimensional histograms produced by it. Let P be the total number of the neighboring pixels, then the LBP operator will have 2^P distinct values, resulting in a 2^P-dimensional histogram. For example, the size of the LBP histogram will be 256/65536 if 8/16 neighboring pixels are considered, and will increase rapidly to a huge number if more neighboring pixels are taken into consideration.

Thus, a dimensionality reduction method for the LBP is needed to address this problem. A straightforward way is to only use small number of the neighboring pixels. For example, the LBP with 8 neighboring pixels is the mostly used one in the applications, and it produces a rather long (256-dimensional) histogram, see the left column of Fig. 2 for an illustration. The size of the LBP histogram will significantly reduce to 16 if only 4 neighboring pixels are taken into account, as illustrated in the middle column of Fig. 2. However, doing this also decreases the discriminative power of the LBP because compared with 8 neighbors, only horizontal and vertical neighbors are considered, and the information of diagonal neighborhood is discarded. Therefore, we use another LBP operator with 4 neighbors to encode the diagonal information, and then combine it with the first one, as shown in Fig. 2. This leads to a final LBP histogram of 32 dimensions, which is much more compact than the original one (256 dimensions). Meanwhile, this combination could keep the discriminative power of the original LBP because it reserves the same number of distinct binary patterns ($2^4 \times 2^4$) as before (2^8).

Since the neighboring pixels used in each unit LBP operator are orthogonal in position, we denote this method as orthogonal local binary patterns combination (OLBPC). It could be generalized to the LBP operators with more neighboring pixels, and the general process is as follows. The neighboring pixels of the original LBP is firstly split into several non-overlapped orthogonal groups, then the LBP code is computed separately for each group, and finally these codes are concatenated together as the new LBP code.

2.3. Comparison with other popular LBP dimensionality reduction methods

We compare our method here with other two popular LBP dimensionality reduction methods — the “uniform patterns” [19] and the CS-LBP [25] on operator level, in terms of discriminative power and feature dimension. The comparisons in the context of local region descriptor will be presented in section 5. In [25], the authors proposed center-symmetric local binary pattern (CS-LBP) for dimensionality reduction. They modified the scheme of how to compare the pixels in the neighborhood. Instead of comparing each pixel with the central pixel, they compare center-symmetric pairs of pixels. This could halve the number of comparisons compared with the original LBP. In [19], the authors proposed the concept of “uniform patterns”, which are certain parts of the original LBP, and are considered to be the fundamental properties of texture. These patterns are called “uniform” because they have one thing in common: no more than two spatial transitions (one-to-zero or zero-to-one) in the circular binary code. For P neighboring pixels, they lead to a histogram of $P \times (P-1) + 3$ dimensions. The “uniform patterns” have been proven to be an effective way for LBP dimensionality reduction [26].

Table 1 Comparison of different dimensionality reduction methods for the LBP (P,R — P neighboring pixels equally located on a circle of radius R)

P,R	**LBP**		**Uniform patterns**		**CS-LBP**		**OLBPC**	
	Bins	**Result**	**Bins**	**Result**	**Bins**	**Result**	**Bins**	**Result**
4,1	16	58.5%	15	58.8%	4	27.8%	16	58.5%
8,1	256	61.4%	59	66.1%	16	50.2%	32	65.4%
12,2	4096	68.7%	135	72.4%	64	61.8%	48	72.7%
16,2	65536	67.6%	243	73.4%	256	54.7%	64	73.2%
20,3	1048576	–	383	74.0%	1024	55.7%	80	74.6%

Since the LBP operator is originally designed as a texture feature, a standard texture classification dataset [27] is chosen to carry out the comparisons. This dataset, namely Outex_TC_00014, contains images of 68 different textures, such as canvas, carpet, granite, tile, sandpaper, wood, and so on. Each kind of texture produces three images of size 746×538 pixels under three different illuminants: 2856K incandescent CIE A light source (Inca), 2300K horizon sunlight (Horizon) and 4000K fluorescent TL84 (TL84). Then each image is equally divided into 20 non-overlapping sub-images of size 128×128 pixels, resulting in 1360 images for each illuminant. The training set is constituted by half of the images under the Inca illuminant, and the test set is constituted by half of the images under the two other illuminants (Horizon and TL84). Therefore, the total numbers of training and test images are 680 and 1360 respectively.

For texture classification, we follow the same process for all the operators — the original LBP, “uniform patterns”, CS-LBP, and our OLBPC. For each image in the training/test set, the operator is applied on all the pixels of the image to get their binary pattern values, and the histogram computed throughout the image is then used as its texture feature. The support vector machine (SVM) algorithm is applied for classification. We compute the χ^2 distance as equation (2) to measure the similarity between each pair of the feature vectors F and F’ (n is the size of both feature vectors):

$$dist_{\chi^2}(F,F') = \sum_{i=1}^{n} \frac{(F_i - F_i')^2}{F_i + F_i'} \tag{2}$$

Then, the kernel based on this distance as equation (3) is used for SVM:

$$K_{\chi^2}(F,F') = e^{-\frac{1}{D} dist_{\chi^2}(F,F')} \tag{3}$$

where D is the parameter for normalizing the distances. Here D is set to the average value of distance between each pair of images in the training set. Finally, each test image is classified into texture category with maximum SVM output decision value. We tune the parameters of the classifier on the training set via cross-validation, and get classification results on the test set.

The classification results and comparisons are presented in Table 1. It can be seen that the classification accuracy generally keeps improving when the number of the neighboring pixels increases, indicating that more neighbors indeed benefit the operator’s performance. However, the increment speed of histogram size for the original LBP is devastating. For example, the histogram size of the LBP with 20 neighboring pixels is so enormous that it is impractical to be used directly. This shows the importance of dimensionality reduction for the LBP. The CS-LBP operator reduces the histogram size of the LBP to its square root, but it also decreases the classification performance. One possible reason is that it discards the information of central pixel in comparison. The “uniform patterns” show good performances, because it significantly reduces the histogram size of the LBP, while still keeping high discriminative power. Actually, it performs even a little better than the original LBP, because it only keeps the most important part of the LBP and removes the other disturbances. Compared with above two methods, our OLBPC operator is more efficient, because it outperforms the CS-LBP and gets almost the same high performance as the “uniform patterns” with the smallest histogram size among them. So it is more suitable for local image region description.

3. Local region description with OLBPC

We construct a new local region descriptor based on the OLBPC operator by following the way similar to the SIFT and CS-LBP descriptors. Fig. 3 depicts the construction process. The input of the descriptor is a normalized local image region around the keypoint, which is either detected by certain interest point detector such as Harris-Laplace, or located on a dense sampling grid. The OLBPC operator is then applied on all the pixels in the region to get their binary

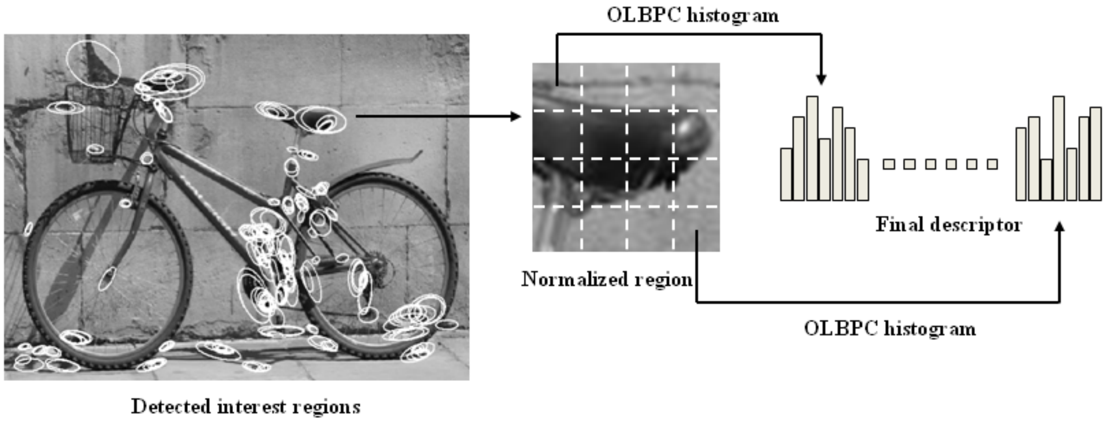

Fig. 3 Construction of local image descriptor with OLBPC

pattern values. In order to include coarse spatial information, the region is equally divided into several small cells, within which a histogram is built based on the binary pattern values of all the pixels. The final descriptor is constructed by concatenating all the histograms from each cell. We adopt the uniform strategy for pixel weighting, as CS-LBP descriptor, and a SIFT-like approach for descriptor normalization. The descriptor is firstly normalized to unit length, each value is then restricted to be no larger than 0.2 (threshold) so that the influence of very large values is reduced, and finally the descriptor is renormalized to unit length.

4. Color OLBPC descriptors

In order to incorporate color information into the OLBPC descriptor to increase its discriminative power, as well as to increase its photometric invariance properties of dealing with different kinds of illumination changes (as described in section 4.1), we extend the OLBPC descriptor to different color space and propose several color OLBPC descriptors in this paper.

4.1. Model analysis for illumination changes

Changes in illumination can be expressed by the diagonal model as equation (4) and the diagonal-offset model as equation (5), where u and c represent respectively the values before and after illumination transformation:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} \tag{5}$$

Based on these two models, different kinds of illumination changes can be expressed as follows [17]:

Light intensity change Image values change by a constant factor in all channels (a = b = c):

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \tag{6}$$

Light intensity shift Image values change by an equal offset in all channels ($a = b = c = 1$, $O_1 = O_2 = O_3$):

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} O_1 \\ O_1 \\ O_1 \end{pmatrix} \tag{7}$$

Light intensity change and shift Image values change by combining two kinds of change above:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} O_1 \\ O_1 \\ O_1 \end{pmatrix} \tag{8}$$

Light color change Image values change in all channels independently ($a \neq b \neq c$), as equation (4).

Light color change and shift Image values change in all channels independently with arbitrary offsets ($a \neq b \neq c$ and $O_1 \neq O_2 \neq O_3$), as equation (5).

4.2. Color OLBPC descriptors and their properties

Six color OLBPC descriptors are proposed in this paper. The main idea is calculating the OLBPC descriptor independently over all the channels of certain color space, and then concatenating them to get the final color OLBPC descriptor, as shown in Fig. 4.

The RGB, HSV, and OPPONENT color spaces are chosen for calculating the color OLBPC descriptors because of their own characteristics. RGB is the most popular color space used in electronic systems for sensing, representation and display of images. It uses additive color mixing with primary colors of red, green and blue to reproduce a broad array of colors. HSV color space rearranges the geometry of RGB so that it could be more relevant to human perception, because it is more natural to think about a color in terms of hue and saturation than in terms of additive color components. OPPONENT color space is constructed to be consistent with human visual system, because it is proven more efficient for human visual system to record differences between responses of cones, rather than each type of cone's individual response. Details of the proposed color OLBPC descriptors and their properties are as follows:

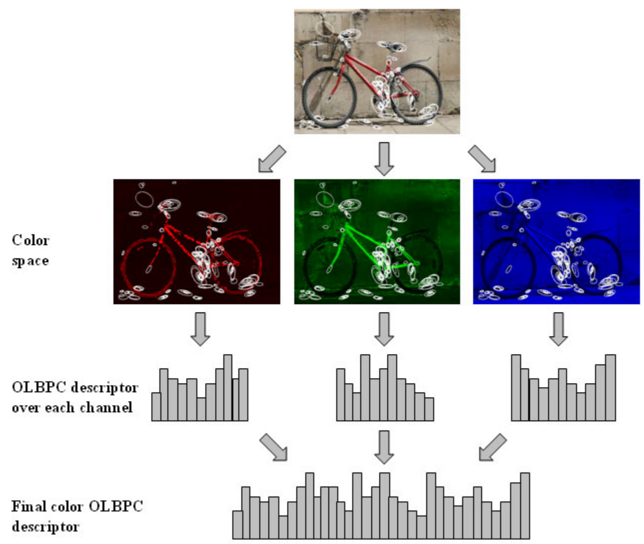

Fig. 4 Calculation of the color OLBPC descriptor

RGB-OLBPC This color descriptor is obtained by computing the OLBPC descriptor over all three channels of the RGB color space. It is invariant to monotonic light intensity change due to the property of the original OLBPC descriptor.

NRGB-OLBPC This color descriptor is obtained by computing the OLBPC descriptor over both r and g channels of the normalized RGB color space as equation (9) (b channel is redundant because r + g + b = 1):

$$\begin{pmatrix} r \\ g \end{pmatrix} = \begin{pmatrix} R/(R+G+B) \\ G/(R+G+B) \end{pmatrix} \tag{9}$$

Due to the normalization, the change factors can be cancelled out if they are constant in all channels. This is proven as equation (10) (Let *a* be the constant factor):

$$\begin{aligned} \begin{pmatrix} r \\ g \end{pmatrix} &= \begin{pmatrix} R/(R+G+B) \\ G/(R+G+B) \end{pmatrix} = \begin{pmatrix} aR'/(aR'+aG'+aB') \\ aG'/(aR'+aG'+aB') \end{pmatrix} \\ &= \begin{pmatrix} aR'/a(R'+G'+B') \\ aG'/a(R'+G'+B') \end{pmatrix} = \begin{pmatrix} R'/(R'+G'+B') \\ G'/(R'+G'+B') \end{pmatrix} \end{aligned} \tag{10}$$

Therefore, r and g channels are scale-invariant, which make this descriptor invariant to light intensity change as equation (6).

OPPONENT-OLBPC This color descriptor is obtained by computing the OLBPC descriptor over all three channels of the OPPONENT color space as equation (11):

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \\ (R+G+B)/\sqrt{3} \end{pmatrix} \tag{11}$$

Due to the subtraction in O_1 and O_2, the change offsets can be cancelled out if they are equal in all channels. This is proven as equation (12) (Let *a* be the equal offset):

$$\begin{aligned} \begin{pmatrix} O_1 \\ O_2 \end{pmatrix} &= \begin{pmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \end{pmatrix} \\ &= \begin{pmatrix} ((R'+a)-(G'+a))/\sqrt{2} \\ ((R'+a)+(G'+a)-2(B'+a))/\sqrt{6} \end{pmatrix} \\ &= \begin{pmatrix} (R'-G')/\sqrt{2} \\ (R'+G'-2B')/\sqrt{6} \end{pmatrix} \end{aligned} \tag{12}$$

Therefore, O_1 and O_2 channels are invariant to light intensity shift as equation (7). O_3 channel represents the intensity information, and has no invariance properties.

NOPPONENT-OLBPC This color descriptor is obtained by computing the OLBPC descriptor over two channels of the normalized OPPONENT color space as equation (13):

$$\begin{pmatrix} O_1' \\ O_2' \end{pmatrix} = \begin{pmatrix} \frac{O_1}{O_3} \\ \frac{O_2}{O_3} \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}(R-G)}{\sqrt{2}(R+G+B)} \\ \frac{R+G-2B}{\sqrt{2}(R+G+B)} \end{pmatrix} \tag{13}$$

Due to the normalization by intensity channel O_3, O_1' and O_2' channels are scale-invariant, which make this descriptor invariant to light intensity change as equation (6).

Hue-OLBPC This color descriptor is obtained by computing the OLBPC descriptor over the Hue channel of the HSV color space as equation (14):

$$Hue = \arctan(\frac{O_1}{O_2}) = \arctan(\frac{\sqrt{3}(R-G)}{R+G-2B}) \tag{14}$$

Due to the subtraction and the division, Hue channel is scale-invariant and shift-invariant, therefore this descriptor is invariant to light intensity change and shift as equation (8).

TC-OLBPC This color descriptor is obtained by computing the OLBPC descriptor over all three channels of the transformed color space as equation (15) (μ is the mean and σ is the standard deviation of each channel):

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} (R-\mu_R)/\sigma_R \\ (G-\mu_G)/\sigma_G \\ (B-\mu_B)/\sigma_B \end{pmatrix} \tag{15}$$

Due to the subtraction and the normalization, all three channels are scale-invariant and shift-invariant, which make this descriptor invariant to light intensity change and shift as equation (8). Furthermore, because each channel is operated independently, this descriptor is also invariant to light color change and shift as equation (5).

It should be noticed that this descriptor has equal values to the RGB-OLBPC descriptor. Because the LBP is computed by taking the subtraction of the neighboring pixels and the central one, the subtraction of the means in this color space is redundant, as this offset is already cancelled out when computing the LBP. And since the descriptor normalization for each channel is done separately, the division of the standard deviation is also redundant. Therefore, the RGB-OLBPC descriptor is used in this paper to represent both descriptors.

Fig. 5 Example image pairs of the Oxford dataset

5. Experimental evaluation

We evaluate the proposed color OLBPC descriptors in three different applications — image matching, object recognition and scene classification. The proposed descriptors are compared with the state-of-the-art descriptors including CS-LBP [25], SIFT [5] and color SIFT [17].

5.1. Experiments on image matching

We adopt the standard dataset available on the Oxford University website [28] to evaluate the proposed descriptors in the application of image matching. The dataset contains image pairs with different geometric and photometric transformations (image blur, viewpoint change, etc.) and for different scene types (structured and textured). The example image pairs are shown in Fig. 5.

The performances of the descriptors are evaluated by the matching criterion, which is based on the number of correctly and falsely matched regions between a pair of images. Two image regions are considered to be matched if the Euclidean distance between their descriptors is below a threshold. The number of correct matches is determined by the "overlap error" [29]. A match is assumed to be correct if this error value is smaller than 0.5. The results are presented by recall versus 1-presicion curve:

$$\text{recall} = \frac{\#\text{correct matches}}{\#\text{correspondences}}, \quad 1-\text{precision} = \frac{\#\text{false matches}}{\#\text{all matches}} \quad (16)$$

where #correspondences is the ground truth number of matches between the images. By changing the distance threshold, we can obtain the recall versus 1-precision curve.

5.1.1. Experimental setup

To compute the descriptors, an interest region detector is required at first to detect the interest regions in each image. We apply the Harris-Affine detector to detect the corner-like structures in images. It originally outputs the elliptic regions of varying scales, and all the regions are then normalized and mapped to a circular region with fixed radius to obtain scale and affine invariance. The normalized regions are also rotated to the direction of their dominant gradient orientations to obtain the rotation invariance.

We implement the CS-LBP descriptor according to [25]. For interest region detection, region normalization, and the SIFT descriptor computation, we use the software package available on the same website as the dataset [28]. We use the "Color Descriptors" software [30] to extract the color SIFT descriptors.

5.1.2. Parameter selection

There are three parameters need to be selected for the proposed color OLBPC descriptors, including the number of the neighboring pixels for the OLBPC operator (P), the radius of the neighboring circle for the OLBPC operator (R), and the number of the cells for each region (M×M). For simplicity, the parameters P and R are evaluated in pairs, such as (4, 1), (8, 1), (12, 2), (16, 2), (20, 3), etc. Also, we select the parameters based on the OLBPC descriptor, and apply the best settings on all the color OLBPC descriptors.

We evaluate the parameters using the image pair named "Graf" (see Fig. 5). Interest regions are detected from each image respectively, and then matched by applying nearest neighbor strategy. A matching score is obtained by measuring the percentage of the correct matches.

Table 2 Parameter selection results (matching score) for the OLBPC descriptor

Cells / P,R	1×1	2×2	3×3	4×4	5×5
4,1	2.84	19.11	25.43	25.77	25.48
8,1	8.76	26.79	34.07	32.88	31.23
12,2	13.77	33.56	**39.31**	36.75	34.64
16,2	11.43	32.48	38.74	35.67	33.56
20,3	13.03	34.47	38.91	37.26	34.41

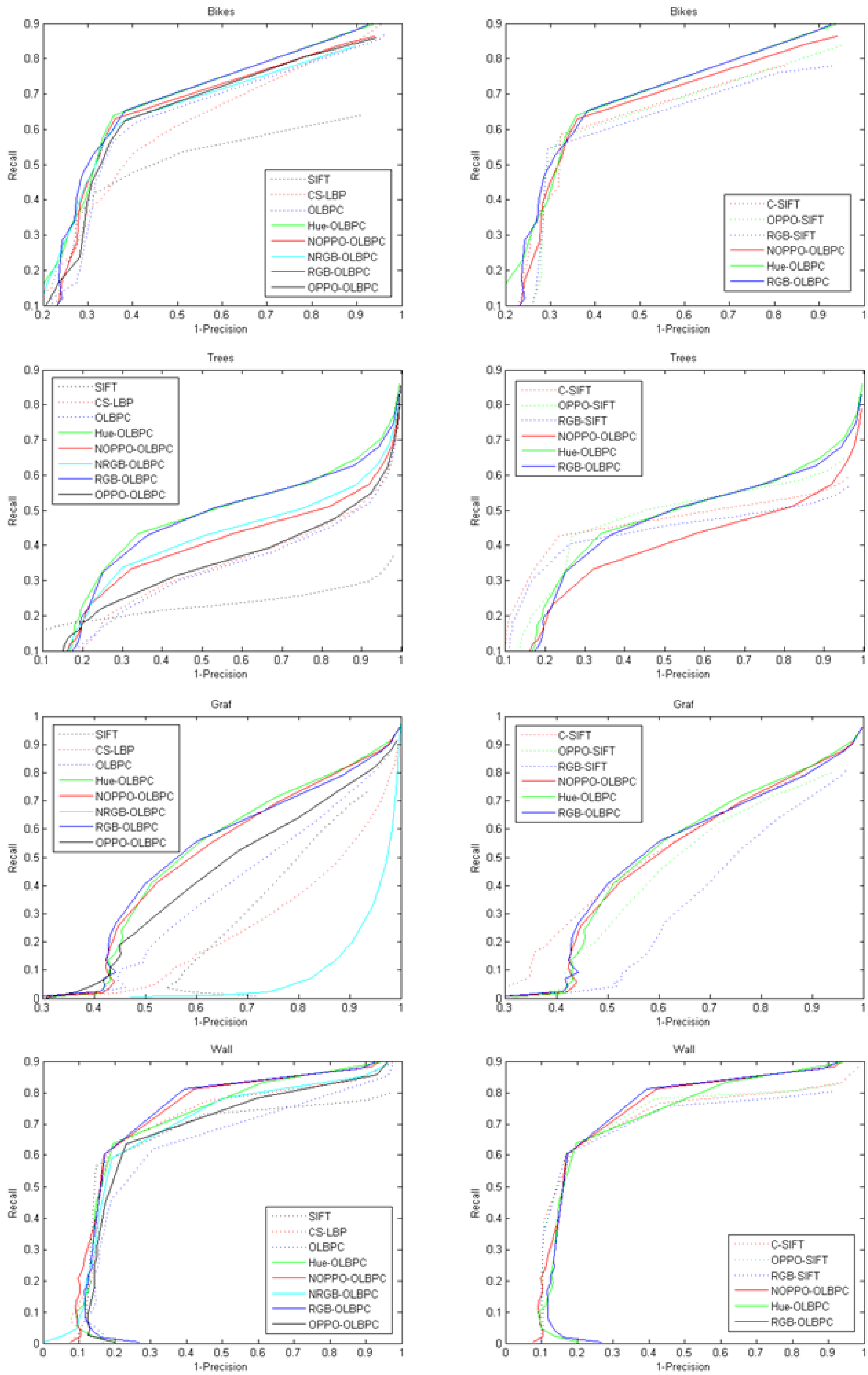

Fig. 6 Image matching results on the Oxford dataset

From the results shown in Table 2, It can be seen that the best performance is obtained when the value of (P, R) pair is set to (12, 2) and the number of the cells is set to 3×3. We apply this parameter setting for the OLBPC descriptor and all the color OLBPC descriptors in the following experiments.

5.1.3. Experimental results

The image matching results on the Oxford dataset are shown in Fig. 6. The figures in the left column show the comparisons of the proposed color OLBPC descriptors with the popular CS-LBP and SIFT descriptors. The figures in the right column show the comparisons of the best three color OLBPC descriptors with the state-of-the-art color SIFT descriptors.

We can see from the results in the left column that: (1) most of the proposed color OLBPC descriptors perform clearly better than the popular CS-LBP and SIFT descriptors; (2) the color OLBPC descriptors outperform the intensity-based OLBPC descriptor in most of the cases, proving the usefulness of introducing color information and additional photometric invariance properties; (3) among the proposed color OLBPC descriptors, Hue-OLBPC, RGB-OLBPC and NOPPONENT-OLBPC descriptors have the best overall performance.

We then compare the best three color OLBPC descriptors with their counterparts — the state-of-the-art color SIFT descriptors. The best three color SIFT descriptors are chosen according to [17]. The results in the right column show that the color OLBPC descriptors also get somewhat better performances than the color SIFT.

5.2. Experiments on object recognition

In order to evaluate the proposed color OLBPC descriptors in the application of object recognition, two standard image datasets are used — the SIMPLIcity database [31] and the PASCAL VOC 2007 benchmark [32].

The SIMPLIcity database is a subset of COREL image database. It contains totally 1,000 images, which are equally divided into 10 different categories: African people, beach, building, bus, dinosaur, elephant, flower, horse, mountain and food. We randomly choose half of the images for training and the other half for test. The recognition accuracy is used as the evaluation criterion. Some example images are shown in Fig. 7.

The PASCAL VOC 2007 benchmark contains nearly 10,000 images of 20 different object classes, such as bike, car, cat, table, person, sofa, train, etc. Each object class contains different number of images, from hundreds to thousands. The dataset is divided into a predefined training set (2501 images), validation set (2510 images) and test set (4952 images). The mean average precision (MAP) is used as the evaluation criterion. Some example images are shown in Fig. 8.

These two datasets have different characteristics. In the SIMPLIcity database, most images have little or no clutter. The objects tend to be centered in each image. Most objects are presented in a stereotypical pose. In the PASCAL VOC 2007 benchmark, all the images are taken from the real-world scenes, with background clutter, occlusions, various viewpoint changes, pose changes, and lighting condition changes, which increase the difficulties of object recognition in this dataset.

5.2.1. Our approach for object recognition

The block diagram of our approach for visual object recognition is depicted in Fig. 9.

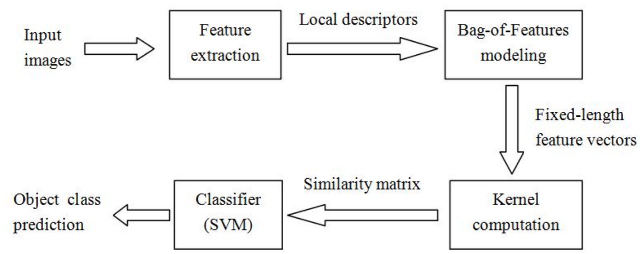

Fig. 9 Flow chart of our approach for object recognition

5.2.2. Feature extraction

We firstly detect the interest points in images by applying the Harris-Laplace salient point detector, which uses a Harris corner detector and subsequently the Laplacian for scale selection. Then a set of local features, including the proposed OLBPC and three best color OLBPC descriptors, CS-LBP, SIFT and three best color SIFT descriptors, are extracted from the local region around each interest point. Unlike the settings in the application of image matching, the descriptors are not rotated to their dominant orientations, because this rotation invariance is useful for image matching, but reduces the accuracy for object recognition.

5.2.3. Bag-of-Features modeling

After the step of feature extraction, each image is represented by a set of local descriptors. The dimensions of these descriptors are still very high because of the large number of the interest points (normally around thousands). Thus, an efficient modeling method is required to transform these high dimensional descriptors to more compact and informative representations.

We apply the popular Bag-of-Features (BoF) method [33] here because of its great success in object recognition tasks. The main idea of the BoF is to represent an image as an orderless collection of local descriptors. More precisely, a visual vocabulary is constructed at first by applying a clustering algorithm on the training data, and each cluster center is considered as a "visual word" in the vocabulary. All the descriptors extracted from an image are then quantized to their closest "visual word" in an appropriate metric space. The number of the descriptors assigned to each "visual word" is accounted into a histogram as the final BoF representation.

Fig. 7 Example images of the SIMPLIcity database

Fig. 8 Example images of the PASCAL VOC 2007 benchmark

Particularly, we build a vocabulary of 1000 "visual words" (for the SIMPLIcity database) or 4000 "visual words" (for the PASCAL VOC 2007 benchmark) for each kind of local feature respectively by applying the k-means clustering algorithm on a subset of the descriptors which are randomly selected from the training data.

5.2.4. Classification

The support vector machine (SVM) algorithm is applied for object classification. Here the LibSVM implementation [34] is used. Once all the local descriptors are transformed to fixed-length feature vectors by the BoF modeling, the χ^2 distance is computed as equation (2) to measure the similarity between each pair of feature vectors. Then, the kernel function based on this distance as equation (3) is used for SVM training and prediction.

For the SIMPLIcity database, each image is classified into the category with maximum SVM output decision value. We tune the parameters of the classifier on the training set via cross-validation, and get the classification results on the test set. For the PASCAL VOC 2007 benchmark, the precision-recall curve is plotted according to the output decision values of SVM classifier, and the MAP is computed based on the proportion of the area under this curve. We train the classifier on the training set, then tune the parameters on the validation set, and get the classification results on the test set.

5.2.5. Experimental results

The object recognition results of the proposed descriptors on the PASCAL VOC 2007 benchmark are shown in Table 3. The comparisons with other state-of-the-art descriptors are also included. It can be seen that: (1) the proposed OLBPC descriptor gets the performance of MAP 38.7%, which is comparable and somewhat better than the popular CS-LBP and SIFT descriptors; (2) the best three color OLBPC descriptors (Hue-OLBPC, NOPPONENT-OLBPC and RGB-OLBPC) get the results of 40.3%, 40.9% and 40.9% respectively, which outperform the intensity-based OLBPC, as well as the CS-LBP and SIFT descriptors for about 2% ~ 3%, indicating that they truly benefit from the additional color information and illumination invariance properties; (3) compared with the state-of-the-art color SIFT descriptors, the best three color OLBPC descriptors get comparable or even slightly better results.

After analyzing the detailed results in Table 3 by each object category, we could observe that the LBP-based descriptors usually perform better on the non-rigid object categories such as bird, cat, dog, horse, person, plant and sofa, while the SIFT-based descriptors are usually better for the rigid object categories such as bicycle, bottle, chair, table, motor, train and monitor. Also, the color descriptors with different photometric invariance properties perform differently on the same object category. Therefore, we

Table 3 Object recognition results on the PASCAL VOC 2007 benchmark

AP (%)	OLBPC	Hue-OLBPC	NOPPONENT-OLBPC	RGB-OLBPC	CS-LBP	SIFT	OPPONENT-SIFT	C-SIFT	RGB-SIFT
airplane	62.2	64.3	64.2	61.9	59.2	56.0	59.9	58.7	57.8
bicycle	38.6	35.4	39.1	42.0	44.8	44.9	43.8	38.9	44.6
bird	25.9	32.9	34.8	32.1	27.4	28.2	27.7	32.1	22.5
boat	56.4	56.0	60.8	59.5	53.0	45.7	49.1	51.8	46.6
bottle	15.0	20.4	20.0	20.3	19.5	19.6	21.2	21.4	21.0
bus	37.8	35.5	35.0	41.1	33.2	37.7	38.0	32.5	37.7
car	62.6	60.5	61.4	65.1	63.1	55.0	57.4	53.2	56.1
cat	38.9	39.3	39.7	42.9	40.2	36.5	37.7	34.1	37.3
chair	39.0	40.5	41.3	39.3	38.7	44.5	42.4	45.9	43.5
cow	20.6	21.5	14.6	24.9	18.3	25.9	17.0	16.6	27.8
table	35.0	36.1	37.0	32.0	33.1	29.6	36.7	38.7	29.1
dog	32.8	35.3	29.4	33.4	31.7	26.5	29.8	29.1	28.8
horse	57.6	64.6	63.6	58.3	55.2	57.0	59.1	61.9	54.8
motor	36.9	39.2	41.7	37.3	34.1	30.2	33.9	44.4	32.1
person	74.1	77.2	75.5	74.7	73.0	73.1	74.5	76.6	72.7
plant	21.3	22.7	26.7	20.1	17.5	11.5	19.9	27.1	11.5
sheep	12.3	23.5	26.0	19.9	16.9	27.4	31.2	30.9	19.4
sofa	25.8	27.8	27.5	25.0	19.0	23.6	22.9	23.2	24.6
train	56.1	44.2	51.7	55.5	56.8	53.4	54.5	58.5	51.1
monitor	25.6	29.2	27.9	31.8	31.7	33.7	35.0	27.3	35.6
Mean	**38.7**	**40.3**	**40.9**	**40.9**	**38.3**	**38.0**	**39.6**	**40.1**	**37.7**

Table 4 Fusion results of the color OLBPC and the color SIFT on PASCAL VOC 2007

AP (%)	FUSION (3 Color OLBPC)	FUSION (3 Color SIFT)	FUSION (3 Color OLBPC + 3 Color SIFT)
airplane	67.0	61.8	67.8
bicycle	48.0	49.8	56.4
bird	36.7	35.0	43.4
boat	62.2	52.9	60.9
bottle	17.6	23.6	26.2
bus	46.4	44.4	51.3
car	67.8	61.7	68.6
cat	45.8	41.7	46.2
chair	43.6	48.2	48.6
cow	26.9	29.1	29.2
table	43.2	41.8	48.2
dog	35.8	32.9	39.3
horse	64.9	64.8	69.6
motor	46.1	48.3	53.3
person	77.8	77.3	79.2
plant	27.3	26.5	31.3
sheep	24.3	33.8	31.7
sofa	32.4	30.6	37.5
train	60.1	62.9	68.3
monitor	35.1	38.1	39.5
Mean	**45.5**	**45.3**	**49.8**

further combine different color descriptors, as well as the color OLBPC and the color SIFT by late fusion to check if they could provide complementary information to each other. The fusion results are shown in Table 4.

It can be observed that: (1) a great performance improvement (about 5%) can be obtained by fusing different color descriptors, both for the OLBPC and the SIFT, proving that different color descriptors are not entirely redundant; (2)

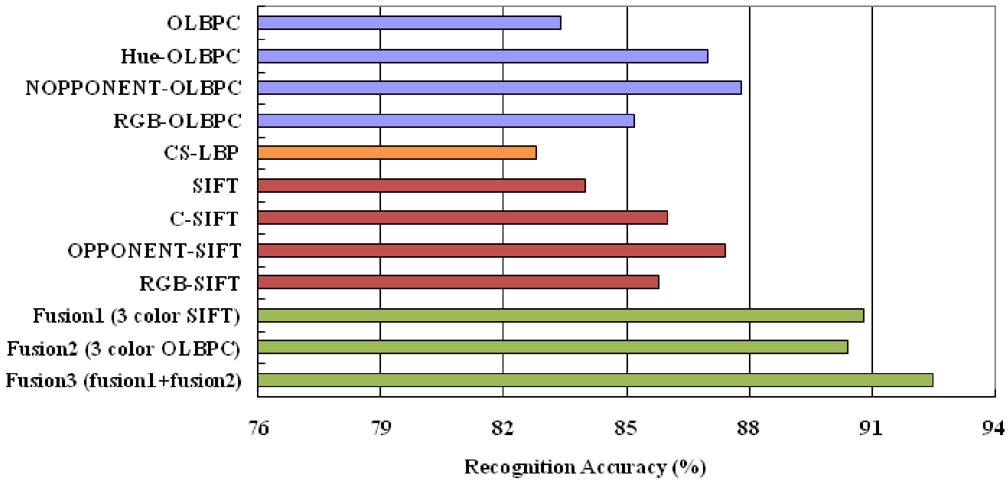

Fig. 10 Object recognition results on the SIMPLIcity database

the color OLBPC descriptors still get comparable or slightly better results than the color SIFT after fusion; (3) the performance can be further improved (more than 4%) by fusing the color OLBPC and the color SIFT descriptors, indicating that these two kinds of descriptors could provide complementary information to each other.

The object recognition results on the SIMPLIcity database are shown in Fig. 10. The similar observations to that on the PASCAL VOC benchmark can be noticed. The color OLBPC descriptors outperform the CS-LBP, SIFT, as well as the intensity-based OLBPC descriptor by about 3% ~ 5% on average, and get comparable results with the color SIFT descriptors. Further improvement (nearly another 5%) can be obtained by fusing three color OLBPC and three color SIFT descriptors, since they provide complementary information to each other.

5.3. Experiments on scene classification

We also evaluate the proposed descriptors in the application of scene classification. The dataset from Oliva and Torralba [35] is used, and denoted as OT scene dataset. It consists of 2688 color images from 8 scene categories: coast (360 samples), forest (328 samples), mountain (374 samples), open country (410 samples), highway (260 samples), inside city (308 samples), tall buildings (356 samples) and streets (292 samples). Fig. 11 shows the example images of each category.

5.3.1. Experimental setup

We mainly follow the same approach described in section 5.2.1 for scene classification. The differences are as follows. Instead of detecting the interest points in images using the Harris-Laplace detectors, we apply the dense sampling strategy to locate the keypoints for local descriptor computation. This is because we prefer to focus on the contents of the whole image, rather than the "object" part only, for scene classification. Particularly, the sampling spacing is set to 6 pixels, resulting in around 1700 keypoints per image. A visual vocabulary of 2000 "visual words" is constructed for each kind of local descriptor to build their Bag-of-Features representations.

We randomly choose half of the images from each scene category for training, and the other half for test. The recognition accuracy is used as the evaluation criterion. We tune the parameters of the classifier on the training set via cross-validation, and get the classification results on the test set.

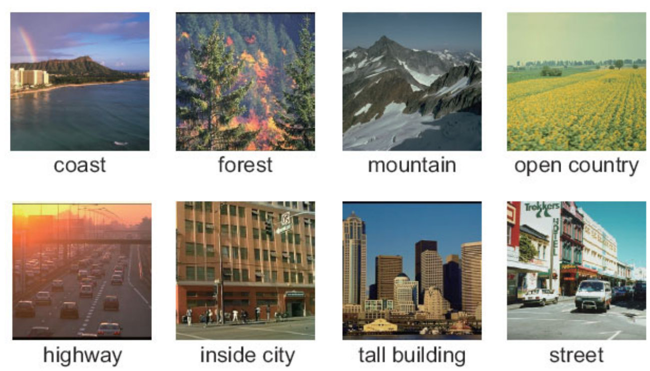

Fig. 11 Example images of the OT scene dataset

5.3.2. Experimental results

The classification results on the OT scene dataset are shown in Fig. 12. The observations are consistent with that in the application of object recognition. The effectiveness of the proposed color OLBPC descriptors are proven by their superior performances to the CS-LBP and SIFT descriptors, and also by their ability of being complementary to the state-of-the-art color SIFT descriptors. It is worthy to be noticed that the NOPPONENT-OLBPC descriptor does not perform well in this case, while its performance is quite good in the

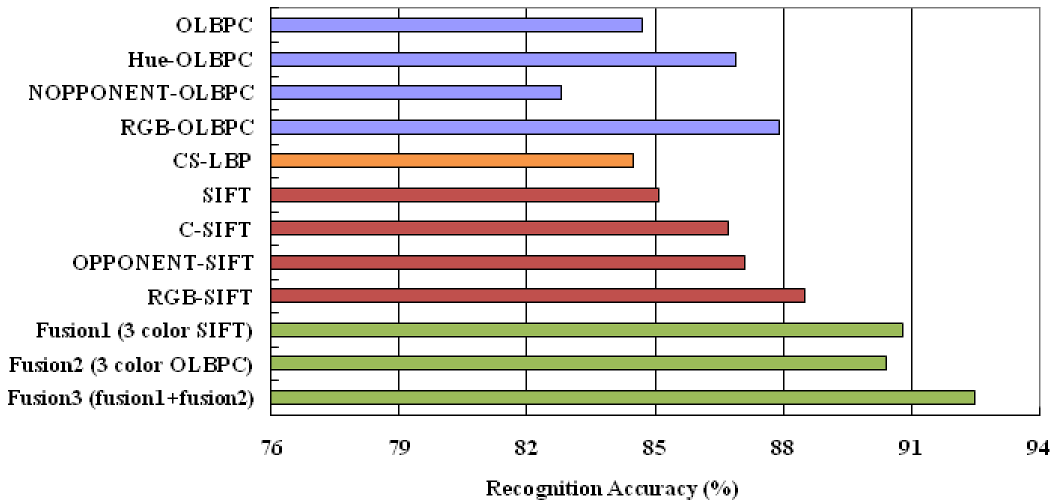

Fig. 12 Classification results on the OT scene dataset

application of object recognition. We believe the main reason is that the OT scene dataset contains more varieties of the illumination changes than the object recognition datasets, and the NOPPONENT-OLBPC descriptor is deficient in power of dealing with these variations, because it is only invariant to light intensity change. This also explains why the RGB-OLBPC and the RGB-SIFT perform the best among the color descriptors, since they possess the strongest invariance properties (invariant to light color change and shift).

5.4. Comparison of the computational cost

As we state in the introduction, a good local descriptor should be both discriminative enough and computationally efficient. The discriminative power of the proposed color OLBPC descriptors has been demonstrated by the previous experiments and applications, and they get comparable or even slightly better performances than the state-of-the-art color SIFT descriptors. Here we show their computational efficiency by comparing them with the color SIFT.

Table 5 Comparison of the computational cost between the color OLBPC and the color SIFT

	Computation Time (s)	MAP (%)
Color OLBPC	1.12	40.9
Color SIFT	2.39	40.1

The comparisons are conducted on the PASCAL VOC 2007 benchmark, and on an Intel Core 2 Duo CPU @ 3.16 GHz with 3GB RAM. We implement the color OLBPC descriptors in C, and use "Color Descriptors" software [30] to compute the color SIFT descriptors. We record the average computation time required for extracting the color OLBPC and color SIFT per image (about size of 500×300) respectively in Table 5. It can be seen that the color OLBPC is about 2 times faster than the color SIFT, and thus is more suitable for large scale problems.

6. Conclusions

In this paper, we firstly proposed a new method namely orthogonal local binary patterns combination (OLBPC) for the dimensionality reduction of the original LBP operator, and then proposed several new local descriptors based on the OLBPC, namely the color OLBPC descriptors, for image region description. The proposed descriptors could incorporate color information to increase their discriminative power, and also to increase their photometric invariance properties of dealing with different illumination changes. The experiments in three different applications showed the effectiveness of the proposed descriptors. They outperform the popular SIFT and CS-LBP descriptors, and get comparable or even slightly better performances than the state-of-the-art color SIFT descriptors. Meanwhile, they could provide complementary information to the color SIFT, since further improvement can be obtained by fusing them. Moreover, the proposed descriptors are about 2 times faster to compute than the color SIFT descriptors. Therefore, they are more promising for large scale problems.

References

[1] M.J. Swain, D.H. Ballard, Color indexing, International Journal of Computer Vision 7 (1) (1991) 11-32.

[2] M.A. Stricker, M. Orengo, Similarity of color images, in Proc. of Storage and Retrieval for Image and Video Databases, 1995, pp. 381-392.

[3] D.K. Park, Y.S. Jeon, C.S. Won, Efficient use of local edge histogram descriptor, in Proc. of ACM workshops on Multimedia, 2000, pp. 51-54.

[4] M. Tuceryan, A.K. Jain, Texture analysis, Handbook of Pattern Recognition and Computer Vision, 2nd edition, World Scientific Publishing Co., 1998, pp. 207-248.

[5] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91-110.

[6] Y. Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 506-513.

[7] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. on Pattern Analysis and Machine Intelligence 27 (10) (2005) 1615-1630.

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

[9] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886-893.

[10] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, International Journal of Computer Vision 73 (2) (2007) 213-238.

[11] J. Li, N.M. Allinson, A comprehensive review of current local features for computer vision, Neurocomputing 71 (2008) 1771-1787.

[12] M. Everingham, Overview and results of the classification challenge, The PASCAL Visual Object Classes Challenge Workshop, 2008.

[13] A.E. Abdel-Hakim, A.A. Farag, CSIFT: a SIFT descriptor with color invariant characteristics, in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1978-1983.

[14] A. Bosch, A. Zisserman, X. Muoz, Scene classification using a hybrid generative / discriminative approach, IEEE Trans. on Pattern Analysis and Machine Intelligence 30 (4) (2008) 712-727.

[15] J. van de Weijer, T. Gevers, A. Bagdanov, Boosting color saliency in image feature detection, IEEE Trans. on Pattern Analysis and Machine Intelligence 28 (1) (2006) 150-156.

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

[17] K.E.A. van de Sande, T. Gevers, C.G.M. Snoek, Evaluating color descriptors for object and scene recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence 32 (9) (2010) 1582-1596.

[18] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on feature distribution, Pattern Recognition 29 (1996) 51-59.

[19] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971-987.

[20] T. Maenpaa, T. Ojala, M. Pietikainen, M. Soriano, Robust texture classification by subsets of local binary patterns, in Proc. of International Conference on Pattern Recognition, 2000, pp. 935-938.

[21] T. Maenpaa, M. Pietikainen, T. Ojala, Texture classification by multipredicate local binary pattern operators, in Proc. of International Conference on Pattern Recognition, 2000, pp. 939-942.

[22] T. Ojala, M. Pietikainen, Unsupervised texture segmentation using feature distributions, Pattern Recognition 32 (3) (1999) 477-486.

[23] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence 28 (2006) 2037-2041.

[24] C. Shan, S. Gong, P.W. McOwan, Robust facial expression recognition using local binary patterns, in Proc. of International Conference on Image Processing, 2005, pp. 370-373.

[25] M. Heikkila, M. Pietikainen, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition 42 (3) (2009) 425-436.

[26] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews 41 (4) (2011) 1-17.

[27] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, S. Huovinen, Outex - new framework for empirical evaluation of texture analysis algorithms, in Proc. of International Conference on Pattern Recognition, 2002, pp. 701-706.

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

[29] K. Mikolajczyk, C. Schmid, An affine invariant interest point detector, in Proc. of European Conference on Computer Vision, vol. 1, 2002, pp. 128-142.

[30] http://www.colordescriptors.com/.

[31] J.Z. Wang, J. Li, G. Wiederhold, SIMPLIcity: semantics-sensitive integrated matching for picture libraries, IEEE Trans. on Pattern Analysis and Machine Intelligence 23 (9) (2001) 947-963.

[32] The PASCAL visual object classes (VOC) challenge 2007 homepage, http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/.

[33] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in ECCV Workshop on Statistical Learning in Computer Vision, 2004, pp.1-22.

[34] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[35] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, International Journal of Computer Vision 42 (3) (2001) 145-175.