Pattern Recognition ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Image region description using orthogonal combination of local binary patterns enhanced with color information

Chao Zhu *, Charles-Edmond Bichot, Liming Chen

Université de Lyon, CNRS, Ecole Centrale de Lyon, LIRIS, UMR5205, F-69134, France

ARTICLE INFO

Article history:
Received 2 October 2011
Received in revised form
11 June 2012
Accepted 1 January 2013

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

ABSTRACT

Visual content description is a key issue for machine-based image analysis and understanding. A good visual descriptor should be both discriminative and computationally efficient while possessing some properties of robustness to viewpoint changes and lighting condition variations. In this paper, we propose a new operator called the orthogonal combination of local binary patterns (denoted as OC-LBP) and six new local descriptors based on OC-LBP enhanced with color information for image region description. The aim is to increase both discriminative power and photometric invariance properties of the original LBP operator while keeping its computational efficiency. The experiments in three different applications show that the proposed descriptors outperform the popular SIFT, CS-LBP, HOG and SURF, and achieve comparable or even better performances than the state-of-the-art color SIFT descriptors. Meanwhile, the proposed descriptors provide complementary information to 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 about four times faster to compute than color SIFT.

1. Introduction

Machine-based automatic object recognition and scene classification is one of the most challenging problems in computer vision. The difficulties are mainly due to intra-class variations and inter-class similarities. Therefore, a key issue and the first important step when solving such problems is to generate good visual content descriptions, which should be both discriminative and computationally efficient, while possessing some properties of robustness to changes in viewpoint, scale and lighting conditions.

Early 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 global visual content of an image. While quite efficient to compute, the downside of these global features is their great sensitivity to image variations such as background clutter, occlusion, viewpoint and illumination changes.

For these reasons, global features have gradually given way later on to local image descriptors. Instead of operating on the whole image, the latter is extracted 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. The visual content of an image is then modeled as a Bag-of-Features (BoF) [5] which views an image as an orderless collection of these local descriptors. An image can thus be described by a histogram, using hard or soft assignment over a visual vocabulary of fixed size learnt from a training dataset. Nowadays, the BoF approach built on local image descriptors has become the dominant approach in the field of visual object recognition and has demonstrated its effectiveness in various famous challenges, e.g., PASCAL VOC [6] and ImageCLEF [7].

Many different local image descriptors have been proposed in the literature, and the most famous one is SIFT [8], which is a 3D histogram of gradient locations and orientations. The location is quantized into a 4×4 location grid and the gradient angle is quantized into eight 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, thereby emphasizing the gradients near the region center. Other popular and widely used local descriptors include PCA-SIFT [9], GLOH [10], SURF [11], HOG [12], DAISY [13],

* Corresponding author. Tel.: +33 679056674; fax: +33 472186443.
E-mail addresses: chao.zhu@ec-lyon.fr (C. Zhu), charles-edmond.bichot@ec-lyon.fr (C.-E. Bichot), liming.chen@ec-lyon.fr (L. Chen).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.01.003

Rank-SIFT [14], BRIEF [15], ORB [16], LIOP [17], MROGH [18], and so on. Most of them are related to SIFT, and can be considered as an extension or refinement of the original SIFT.

Several comprehensive studies [10,19,20] have highlighted the interest of local image descriptors in tasks as diverse as image region matching, texture classification, object recognition and scene classification. Among them, SIFT proves 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 [6]. Moreover, since the original SIFT is an intensity based descriptor, several color SIFT descriptors have been proposed [21–24] to capture color information and further enhance its discriminative power. In [25], Van de Sande et al. evaluated different color descriptors in a structured way, and recommended to use color SIFT descriptors for object and scene recognition because they outperform the original SIFT. However, the downside of color SIFT descriptors is their high computational cost, especially when the size of image or the scale of dataset significantly increases. Therefore, it is highly desirable that local image descriptors offer both high discriminative power and computational efficiency.

The local binary pattern (LBP) operator [26] is a well-known texture feature which has been successfully applied to many applications, e.g., texture classification [27–29], texture segmentation [30], face recognition [31,32], and facial expression recognition [33,34]. The LBP operator has several interesting properties. First of all, it is simple and fast to compute. Moreover, it offers strong discriminative power for the description of texture structure while staying robust to monotonic lighting changes. All these advantages make LBP a good candidate for describing local image regions. However, the LBP operator tends to produce high-dimensional feature vectors, especially when the number of considered neighboring pixels increases. The so-called "curse of dimensionality" is a barrier for using it directly as a local region descriptor. Thus, a key issue of making LBP a local region descriptor is to reduce its dimensionality. There exist in the literature two main works, namely "uniform patterns" [27] and center-symmetric local binary pattern (CS-LBP) operator [35], which address this issue.

In this paper, we propose a novel effective dimensionality reduction method for LBP, denoted as the orthogonal combination of local binary patterns (OC-LBP), while keeping high discriminative power of the original LBP in capturing local texture patterns. The basic idea is to first split the neighboring pixels of the original LBP operator into several non-overlapped orthogonal groups, then compute the LBP code separately for each group, and finally concatenate them together. The experimental results on a standard texture classification dataset show that our method is much more effective than both CS-LBP operator and "uniform patterns" in terms of dimensionality reduction, since our method produces the LBP features with the smallest dimensions while still keeping high classification accuracy.

The proposed OC-LBP operator is then adopted to build a distribution-based local image region descriptor, denoted as OC-LBP descriptor, by following a way similar to SIFT: given several local regions of an image, each region is first divided into small cells for spatial information; in each cell, the OC-LBP feature is then computed for each pixel and an LBP histogram is constructed; finally, all the histograms from the cells are concatenated and delivered as the final region descriptor. Our aim is to build a more efficient local descriptor by replacing the costly gradient information with local texture patterns in the SIFT scheme.

It can be noticed that almost all the LBP related descriptors are intensity-based. However, color plays an important role for distinguishing objects and scenes, especially in natural scenes. Meanwhile, there can be various changes in lighting and viewing conditions in real-world scenes, leading to large variations of objects' appearances, thereby further complicating the task of object and scene recognition. Therefore, similar to the extension of SIFT to color SIFT, we further extend the OC-LBP descriptor to different color spaces and propose six color OC-LBP descriptors in this paper to increase the photometric invariance properties and enhance the discriminative power of the intensity-based descriptor. In [36], we have proposed several color LBP features, which are based on the original LBP operator and serve as global features. Different from them, the proposed color OC-LBP descriptors in this paper are based on the orthogonal combination of the LBP operator, and serve as local features. They could thus be considered as the extensions of our previous work [36]. The experimental results in three different applications show that the proposed descriptors outperform the popular SIFT, CS-LBP, HOG and SURF descriptors, and achieve comparable or even better performances than the state-of-the-art color SIFT descriptors. Meanwhile, the proposed descriptors provide complementary information to 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 color SIFT.

The main contributions of this paper are summarized as follows:

(1) We propose a new operator called the orthogonal combination of local binary patterns (denoted as OC-LBP) to reduce the dimensionality of the original LBP operator so that it is feasible to be utilized for image region description.
(2) Based on the proposed OC-LBP operator, we build a new local image region descriptor, denoted as OC-LBP descriptor, with high computational efficiency.
(3) We propose six new color OC-LBP descriptors by extending gray OC-LBP descriptor to different color spaces in order to increase the photometric invariance properties and enhance the discriminative power.

The remaining sections are organized as follows. Section 2 introduces the proposed orthogonal combination of local binary patterns (OC-LBP) in detail, and compares it with other two popular LBP dimensionality reduction methods: "uniform patterns" and CS-LBP operator. The construction of the OC-LBP descriptor for local image regions is described in Section 3. We then give details of the proposed color OC-LBP 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 of LBP

2.1. Original LBP operator

The original LBP operator can be seen as a unified approach to statistical and structural texture analysis. The histogram of the binary patterns computed over a region is generally used for texture description. The LBP operator describes each pixel by the relative gray levels of its neighboring pixels. Fig. 1 illustrates the calculation of the LBP code for one pixel with eight neighbors. 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. Formally, the LBP

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

code of the pixel at (x_c,y_c) is calculated as

$$LBP_{P,R}(x_c,y_c) = \sum_{p=0}^{P-1} S(g_p - g_c) \times 2^p \quad (1)$$

$$S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

where g_c is the value of the central pixel, g_p corresponds to the gray values of the P neighboring pixels equally located on a circle of radius R.

The final LBP feature of an image is generally distribution-based and consists of computing the LBP code for each pixel within the image and building a histogram based on these codes. It can be noticed that the LBP feature is very fast to calculate, and is invariant to monotonic illumination changes. Thus it is a good candidate for local image region description.

However, the drawback of the LBP feature lies in the high dimensionality of histograms produced by the LBP codes. Let P be the total number of neighboring pixels, then the LBP feature will have 2^P distinct values, resulting in a 2^P-dimensional histogram. For example, the size of the LBP histogram will be 256 (65,536, respectively) if 8 (16, respectively) neighboring pixels are considered. It will rapidly increase to a huge number if more neighboring pixels are taken into consideration. Thus, a dimensionality reduction method for LBP is needed to address this problem.

2.2. Orthogonal combination of local binary patterns (OC-LBP)

To reduce the dimensionality of the LBP histogram, a straightforward way is to only consider fewer neighboring pixels. For example, the LBP operator with eight neighbors is mostly used 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 four neighboring pixels are taken into account, as illustrated in the middle column of Fig. 2. However, this brut reduction also decreases the discriminative power of the LBP feature because compared to eight neighbors, only horizontal and vertical neighbors are considered, and the information of diagonal neighborhood is discarded. We need to find out a tradeoff between the reduction of the LBP histogram dimensionality and its descriptive power.

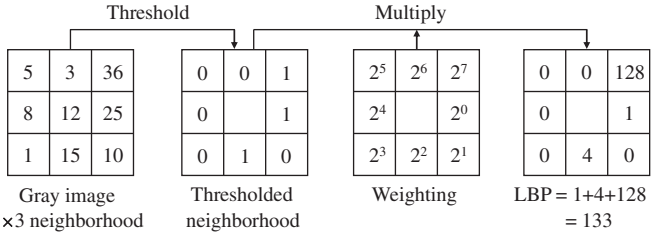

Fig. 1. Calculation of the original LBP code.

In this paper, we propose an orthogonal combination of local binary patterns, namely OC-LBP, which drastically reduces the dimensionality of the original LBP histogram while keeping its discriminative power. Specifically, given P neighboring pixels equally located on a circle of radius R around a central pixel c, OC-LBP is obtained by combining the histograms of $[P/4]$ different 4-orthogonal-neighbor operators, each of which consists of turning the previous four orthogonal neighbors by one position in a clockwise direction. The dimension of an OC-LBP based histogram is thus $2^4 \times [P/4]$ or simply $4 \times P$, which is linear with the number of neighboring pixels in comparison to 2^P for the original LBP-based scheme.

Fig. 2 illustrates the construction process of an OC-LBP operator with eight neighboring pixels. In this case, two regular 4-neighbor LBP operators are considered. The first one consists of the horizontal and vertical neighbors, and the second one consists of the diagonal neighbors. By concatenating these two LBP histograms, we obtain the OC-LBP histogram with 32 dimensions, which is eight times more compact than the original 8-neighbor LBP histogram (256 dimensions). Meanwhile, this combination keeps quite well the discriminative power of the original LBP because it preserves the same number of distinct binary patterns ($2^4 \times 2^4$) as before (2^8).

This orthogonal combination of local binary patterns (OC-LBP) can also be generalized in different ways. For instance, the neighboring pixels of the original LBP can be first split into several non-overlapped orthogonal groups, then the LBP code can be computed separately for each group, and finally the histograms based on these separate LBP codes can be concatenated and used as the image description.

2.3. Comparison with other popular LBP dimensionality reduction methods

In this section, we make a comparison between the proposed OC-LBP and other two popular dimensionality reduction methods

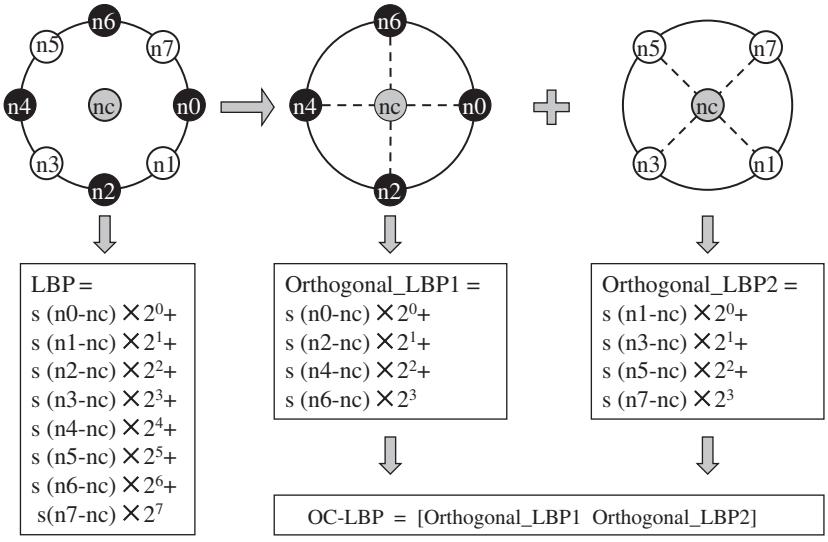

Fig. 2. Calculation of the LBP and OC-LBP operators with eight neighboring pixels.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

for LBP both in terms of discriminative power and feature dimensionality. These two methods, namely "uniform patterns" [27] and CS-LBP [35], are compared here with OC-LBP on operator level. The comparisons in the context of local region descriptor will be presented in Section 5.

In [27], Ojala et al. proposed the concept of "uniform patterns", which are a subset of the original LBP codes, and are considered to convey some fundamental properties of texture. These patterns are called "uniform" because they have one property 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 [37]. In [35], Heikkilä et al. proposed center-symmetric local binary pattern (CS-LBP) operator 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 halves the number of comparisons compared to the original LBP.

Table 1 summarizes the dimensionality of the histograms produced by different methods with P neighboring pixels.

As we can see, the most effective scheme in terms of histogram dimensionality reduction is the proposed OC-LBP, which is linear with P—the number of neighboring pixels, compared to exponential dimension of the original LBP and CS-LBP, and quadratic dimension of "uniform patterns". Then, what about the discriminative power of OC-LBP compared to the original LBP, CS-LBP and "uniform patterns"?

Since the LBP operator is originally designed as a texture feature, a standard texture classification dataset [38] 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: 2856 K incandescent CIE A light source (Inca), 2300 K horizon sunlight (Horizon) and 4000 K 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 features (the original LBP, "uniform patterns", CS-LBP and the proposed OC-LBP). For each image in the training/test set, each of the operators 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) is applied for classification. We compute the χ^2 distance as Eq. (3) to measure the similarity between each pair of the feature vectors F and F' (n is the size of both feature vectors)

$$dist_{\chi^2}(F,F')=\sum_{i=1}^{n}\frac{(F_i-F'_i)^2}{F_i+F'_i} \tag{3}$$

Then, the kernel based on this distance is computed as Eq. (4) for the SVM training and prediction

$$K_{\chi^2}(F,F')=e^{(-1/D)dist_{\chi^2}(F,F')} \tag{4}$$

Table 1
Comparison of the histogram dimensionality of different methods with P neighboring pixels.

LBP	Uniform patterns	CS-LBP	OC-LBP
2^P	$P\times(P-1)+3$	$2^{[P/2]}$	$4\times P$

Table 2
Comparison of different LBP dimensionality reduction methods in terms of histogram size and classification accuracy on Outex_TC_00014 (P,R:P neighboring pixels equally located on a circle of radius R).

P,R	LBP		Uniform patterns		CS-LBP		OC-LBP	
	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	65,536	67.6	243	73.4	256	54.7	64	73.2
20,3	1,048,576	–	383	74.0	1024	55.7	80	74.6

where D is the parameter for normalizing the distances. Here D is set as the average value of distances between each pair of images in the training set. Finally, each test image is classified into texture category with the maximum SVM output decision value. We tune the parameters of the classifier on the training set via 5-fold cross-validation, and obtain the classification results on the test set.

The classification results and comparisons are presented in Table 2. It can be seen that the classification accuracy generally keeps improving when the number of neighboring pixels increases, suggesting that the consideration of more neighbors can be beneficial to the operator's performance. However, the increment speed of histogram size for the original LBP is devastating. For example, the LBP histogram size with 20 neighboring pixels is so enormous that it is impractical to be used directly. This shows the importance of dimensionality reduction for LBP. The CS-LBP operator reduces the LBP histogram size to its square root, but it also decreases the classification accuracy. 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 LBP histogram size, 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 LBP and removes the other disturbances. Compared to these two methods, the proposed OC-LBP operator shows its effectiveness. It outperforms CS-LBP and achieves almost the same high performance as the "uniform patterns" but with the smallest histogram size among them.

3. Local region description with OC-LBP

We construct a new local region descriptor based on the proposed OC-LBP operator by following a way similar to the SIFT [8] and CS-LBP [35] descriptors. Fig. 3 depicts the construction process. The input of the descriptor is a normalized local image region around a keypoint, which is either detected by certain interest point detector such as Harris-Laplace, or located on a dense sampling grid. The OC-LBP operator is then applied on all the pixels in the region to get their binary 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 the cells. We adopt the uniform strategy for pixel weighting, as the CS-LBP descriptor, and a SIFT-like approach for descriptor normalization. The descriptor is first 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. We denote this new local image descriptor as OC-LBP descriptor.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

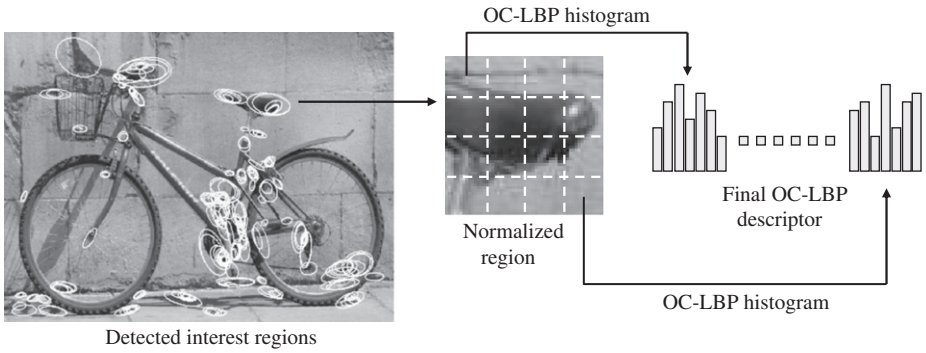

Fig. 3. Construction of local image descriptor with OC-LBP.

4. Color OC-LBP descriptors

The classical LBP-related descriptors only use gray information. However, color information may significantly improve the discriminative power of a descriptor. Moreover, incorporating color information may enhance its photometric invariance properties when dealing with different kinds of illumination changes as described in Section 4.1.

4.1. Model analysis for illumination changes

Changes in illumination can be expressed by the diagonal model as Eq. (5) and the diagonal-offset model as Eq. (6), where u and c represent, respectively, the values before and after illumination transformation

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \tag{5}$$

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} \tag{6}$$

Based on these two models, different kinds of illumination changes can be expressed as follows [25]:

Light intensity change: Image values change by a constant factor in all channels ($\alpha=\beta=\gamma$)

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \tag{7}$$

Light intensity shift: Image values change by an equal offset in all channels ($\alpha=\beta=\gamma=1, O_1=O_2=O_3$)

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

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

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} O_1 \\ O_1 \\ O_1 \end{pmatrix} \tag{9}$$

Light color change: Image values change in all channels independently ($\alpha\neq\beta\neq\gamma$), as Eq. (5).

Light color change and shift: Image values change in all channels independently with arbitrary offsets ($\alpha\neq\beta\neq\gamma$ and $O_1\neq O_2\neq O_3$), as Eq. (6).

4.2. Color OC-LBP descriptors and their properties

In order to incorporate color information, we extend the OC-LBP descriptor to different color spaces and propose six color OC-LBP descriptors in this section. The main idea is to calculate the original OC-LBP descriptor independently over different channels in a given color space, and then concatenate them to get the final color OC-LBP descriptor, as shown in Fig. 4.

The *RGB*, *HSV*, and *OPPONENT* color spaces are chosen for calculating color OC-LBP 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 range 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 proves 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 OC-LBP descriptors and their properties are presented as follows:

RGB-OC-LBP: This color descriptor is obtained by computing the OC-LBP 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 OC-LBP descriptor.

NRGB-OC-LBP: This color descriptor is obtained by computing the OC-LBP descriptor over both r and g channels of the *normalized RGB* color space as Eq. (10) (b channel is redundant because $r+g+b=1$)

$$\begin{pmatrix} r \\ g \end{pmatrix} = \begin{pmatrix} R/(R+G+B) \\ G/(R+G+B) \end{pmatrix} \tag{10}$$

Due to the normalization, the change factors can be canceled out if they are constant in all channels. This is proven as Eq. (11) (let a be the constant factor)

$$\begin{pmatrix} r \\ g \end{pmatrix} = \begin{pmatrix} R/(R+G+B) \\ G/(R+G+B) \end{pmatrix} = \begin{pmatrix} aR'/(aR'+aG'+aB') \\ aG'/(aR'+aG'+aB') \end{pmatrix} = \begin{pmatrix} aR'/a(R'+G'+B') \\ aG'/a(R'+G'+B') \end{pmatrix} = \begin{pmatrix} R'/(R'+G'+B') \\ G'/(R'+G'+B') \end{pmatrix} \tag{11}$$

Therefore, r and g channels are scale-invariant, which makes this descriptor invariant to light intensity change as Eq. (7).

OPPONENT-OC-LBP: This color descriptor is obtained by computing the OC-LBP descriptor over all three channels of the

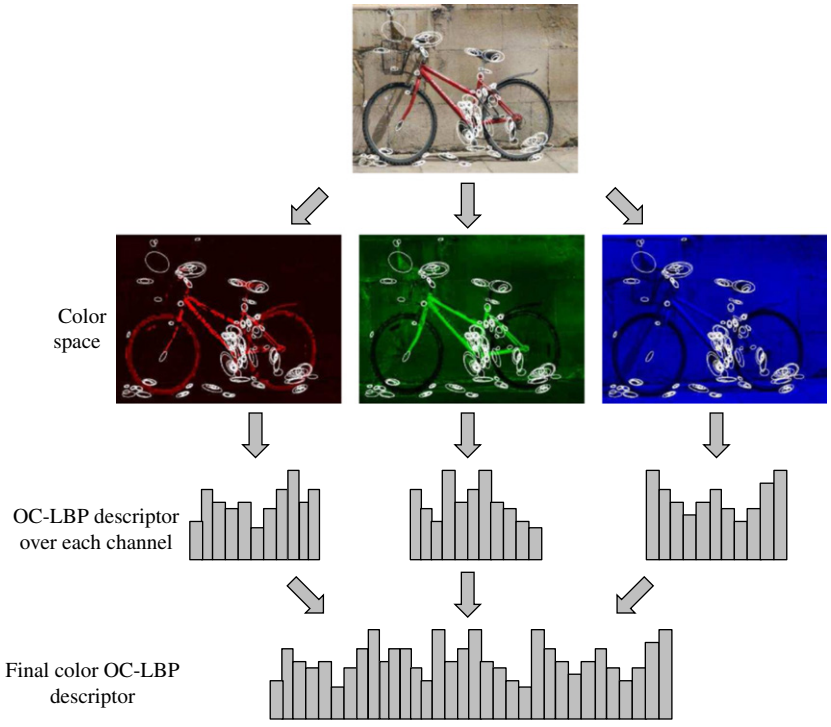

Fig. 4. Calculation of color OC-LBP descriptor.

OPPONENT color space as

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \\ (R+G+B)/\sqrt{3} \end{pmatrix} \tag{12}$$

Due to the subtraction in O_1 and O_2, the change offsets can be canceled out if they are equal in all channels. This is proven as (let a be the equal offset)

$$\begin{aligned} \begin{pmatrix} O_1 \\ O_2 \end{pmatrix} &= \begin{pmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \end{pmatrix} \\ &= \begin{pmatrix} ((R'+a)-(G'+a))/\sqrt{2} \\ ((R'+a)+(G'+a)-2(B'+a))/\sqrt{6} \end{pmatrix} \\ &= \begin{pmatrix} (R'-G')/\sqrt{2} \\ (R'+G'-2B')/\sqrt{6} \end{pmatrix} \end{aligned} \tag{13}$$

Therefore, O_1 and O_2 channels are invariant to light intensity shift as Eq. (8). O_3 channel represents the intensity information, and has no invariance properties.

NOPPONENT-OC-LBP: This color descriptor is obtained by computing the OC-LBP descriptor over two channels of the *normalized OPPONENT* color space as

$$\begin{pmatrix} O'_1 \\ O'_2 \end{pmatrix} = \begin{pmatrix} \dfrac{O_1}{O_3} \\ \dfrac{O_2}{O_3} \end{pmatrix} = \begin{pmatrix} \dfrac{\sqrt{3}(R-G)}{\sqrt{2}(R+G+B)} \\ \dfrac{R+G-2B}{\sqrt{2}(R+G+B)} \end{pmatrix} \tag{14}$$

Due to the normalization by intensity channel O_3, O'_1 and O'_2 channels are scale-invariant, which makes this descriptor invariant to light intensity change as Eq. (7).

Hue-OC-LBP: This color descriptor is obtained by computing the OC-LBP descriptor over the *Hue* channel of the *HSV* color space as

$$Hue = \arctan\left(\frac{O_1}{O_2}\right) = \arctan\left(\frac{\sqrt{3}(R-G)}{R+G-2B}\right) \tag{15}$$

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 Eq. (9).

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

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} (R-\mu_R)/\sigma_R \\ (G-\mu_G)/\sigma_G \\ (B-\mu_B)/\sigma_B \end{pmatrix} \tag{16}$$

Due to the subtraction and the normalization, all three channels are scale-invariant and shift-invariant, which makes this descriptor invariant to light intensity change and shift as Eq. (9). Furthermore, because each channel is operated independently, this descriptor is also invariant to light color change and shift as Eq. (6).

It should be noticed that this descriptor has equal values to the RGB-OC-LBP 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 canceled 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-OC-LBP descriptor is used in this paper to represent both descriptors.

5. Experimental evaluation

We evaluated the proposed intensity-based and color OC-LBP descriptors in three different applications: (1) image matching,

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

(2) object recognition and (3) scene classification. The Oxford image matching dataset [39] is used for parameter selection and image matching experiments. Two standard image datasets – the PASCAL VOC 2007 benchmark [40] and the SIMPLIcity dataset [41] – are applied for object recognition experiments. The scene dataset from Oliva and Torralba [42] is adopted for scene classification experiments. The proposed descriptors are compared with several state-of-the-art descriptors including SIFT [8], color SIFT [25], CS-LBP [35], SURF [11], HOG [12] and GIST [42]. These descriptors have been chosen for their diversity in terms of local visual content characterization. While SIFT and color SIFT are the most popular and successful local descriptors in the literature, HOG is also a popular descriptor which captures local object appearance and shape through the distribution of intensity gradients. As such it is widely used for object detection and recognition. GIST is a popular holistic feature which estimates the dominant spatial structure of a scene to capture a set of perceptual dimensions (naturalness, openness, roughness, expansion and ruggedness). As such it is widely applied for scene classification. SURF is a typical local descriptor using Haar wavelets as features. Finally, CS-LBP is also binary-pattern-based and provides a way for LBP dimensionality reduction, as we introduced in Section 2.3.

5.1. Parameter selection

There are three parameters to be fixed for the proposed OC-LBP descriptors, including the number of neighboring pixels for the OC-LBP operator (P), the radius of neighboring circle for the OC-LBP operator (R), and the number of cells for each region ($M \times M$). 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 gray OC-LBP descriptor, and apply the best settings on all color OC-LBP descriptors.

We adopt the standard Oxford image matching dataset [39] for parameter selection. This dataset contains image pairs with different geometric and photometric transformations (image blur, viewpoint change, illumination change, etc.) and different scene types (structured and textured). The sample image pairs are shown in Fig. 5. Here the image pair named "Graf" is used for parameter selection as in [35]. To compute the descriptors, an interest region detector is required at first to detect 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 use the software package available on the same website as the dataset for interest region detection and normalization. Each detected region is normalized to the size of 41×41 pixels. Then, all the regions from each image are described by the OC-LBP descriptor and are matched according to their Euclidean distances. An ROC curve is obtained by changing the distance threshold for matching, and a matching score is also obtained at the same time by applying nearest neighbor matching strategy and measuring the percentage of the correct matches.

The results are shown in Fig. 6. From sub-figures (1)–(5), it can be seen that 3×3 cells obtain the best results for almost all the pairs of (P, R), both in ROC curve and matching score cases. Therefore, we fix the number of cells to 3×3, and then compare the different values of (P, R) pair in sub-figure (6). The best performance is achieved when $P=12$, $R=2$. We thus apply this parameter setting on gray OC-LBP descriptor and all color OC-LBP descriptors in the following experiments.

5.2. Experiments on image matching

We adopt the same dataset introduced in Section 5.1 to evaluate the proposed descriptors in the application of image matching. 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" [43]. A match is assumed to be correct if this error value is smaller than 0.5. The results are presented by recall versus 1-precision curve

$$recall = \frac{\#correct\ matches}{\#correspondences} \tag{17}$$

$$1\text{-}precision = \frac{\#false\ matches}{\#all\ matches} \tag{18}$$

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.

5.2.1. Experimental setup

We use the software package mentioned in Section 5.1 for interest region detection, region normalization, and SIFT computation. We implement the CS-LBP descriptor according to [35], and apply the same parameter setting as the OC-LBP descriptor for fair comparison. To compute color SIFT descriptors, we use the "ColorDescriptor" software available online [44].

5.2.2. Experimental results

The image matching results on the Oxford dataset are shown in Figs. 7 and 8. Fig. 7 shows the comparisons of the proposed gray

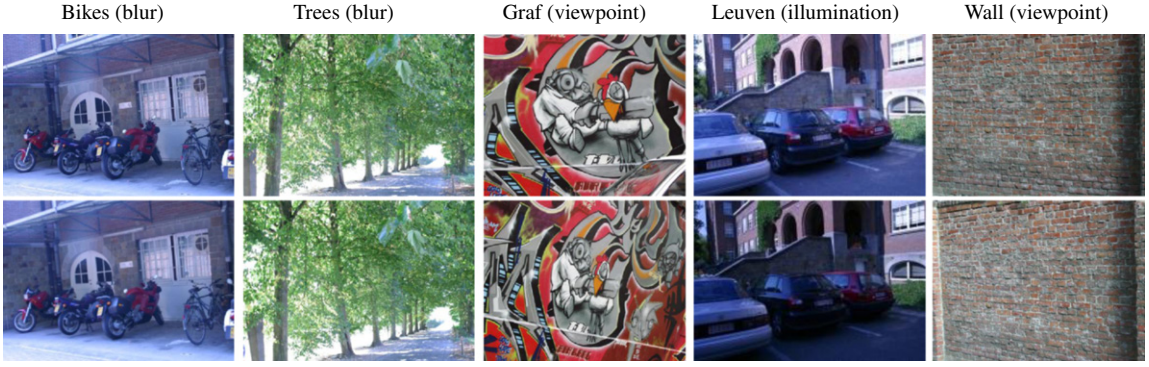

Fig. 5. Sample image pairs of the Oxford image matching dataset.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

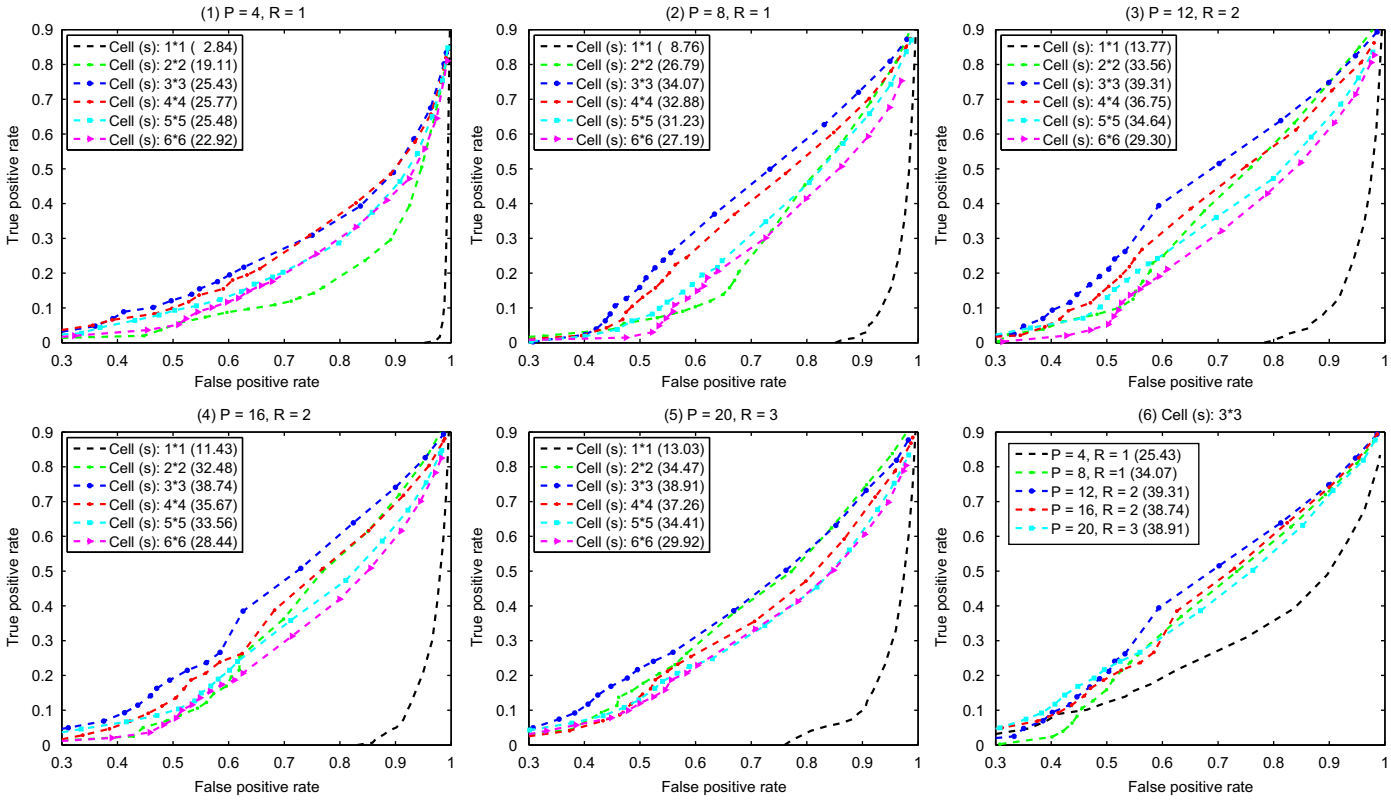

Fig. 6. Parameter selection results for the OC-LBP descriptor (matching score presented between parentheses).

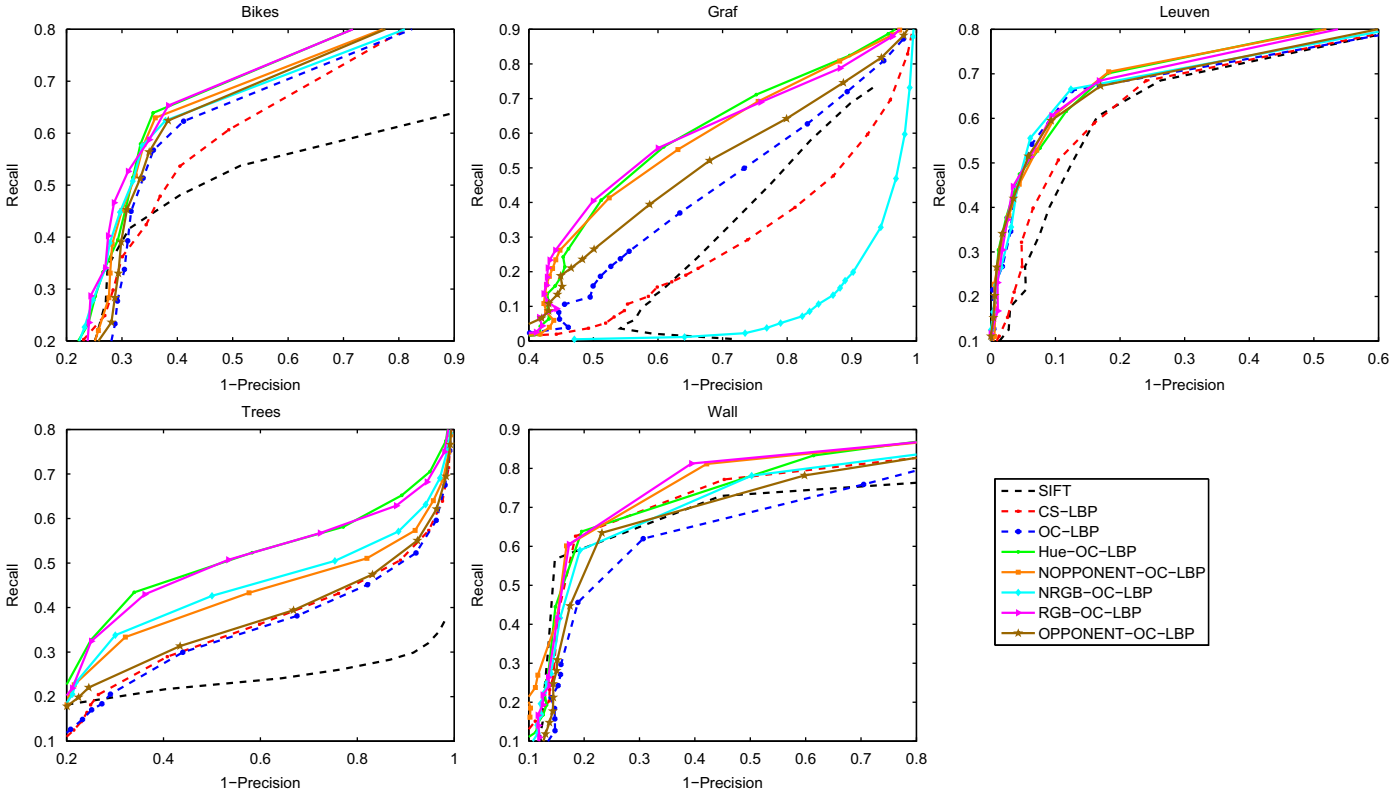

Fig. 7. Image matching results on the Oxford dataset (comparisons of the proposed descriptors with the popular SIFT and CS-LBP descriptors).

and color OC-LBP descriptors with the popular SIFT and CS-LBP descriptors. Fig. 8 shows the comparisons of the best three color OC-LBP descriptors with the state-of-the-art color SIFT descriptors.

We can see from the results in Fig. 7 that: (1) the OC-LBP descriptor performs better than the popular CS-LBP and SIFT descriptors; (2) the color OC-LBP descriptors outperform the

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

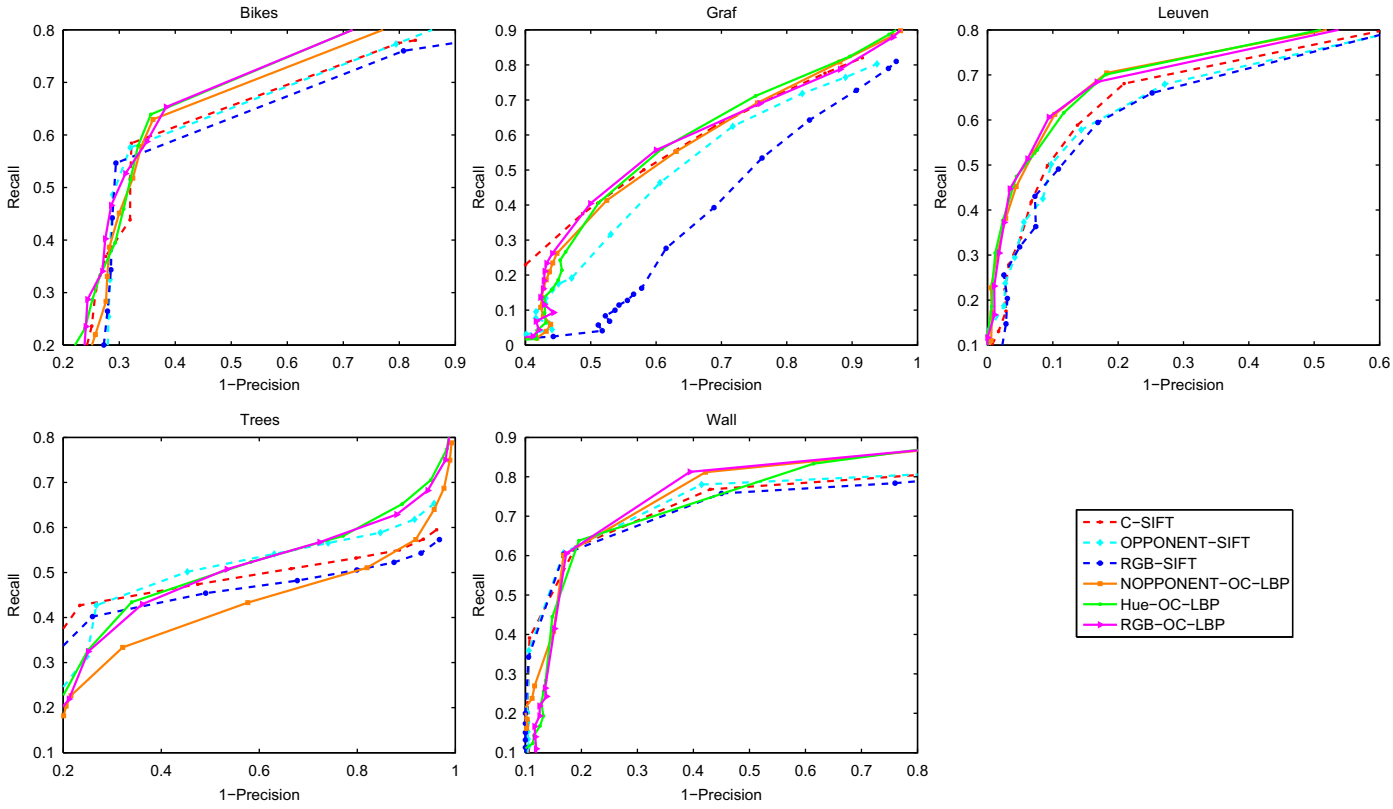

Fig. 8. Image matching results on the Oxford dataset (comparisons of the best three color OC-LBP descriptors with the state-of-the-art color SIFT descriptors). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

intensity-based OC-LBP descriptor in most of the cases, proving the usefulness of incorporating color information and additional photometric invariance properties; (3) among the proposed color OC-LBP descriptors, Hue-OC-LBP, RGB-OC-LBP and NOPPONENT-OC-LBP descriptors have the best overall performance, which is consistent with their strong properties of illumination invariance.

We then compare the best three color OC-LBP descriptors with their counterparts: the state-of-the-art color SIFT descriptors. The best three color SIFT descriptors are chosen according to [25]. The results in Fig. 8 show that the color OC-LBP descriptors also achieve slightly better performances than color SIFT.

5.3. Experiments on object recognition

In order to evaluate the proposed descriptors in the application of object recognition, two standard image datasets are used: the PASCAL VOC 2007 benchmark [40] and the SIMPLIcity dataset [41].

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. 9.

The SIMPLIcity dataset is a subset of COREL image database. It contains totally 1000 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. 10.

These two datasets have different characteristics. In the SIMPLIcity dataset, 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, thus with background clutter, occlusions, and various variations in viewpoint, pose and lighting condition, which increase the difficulties of object recognition in this dataset.

5.3.1. Our approach for object recognition

The block diagram of our approach for visual object recognition is depicted in Fig. 11.

5.3.2. Feature extraction

The interest points in images are first detected by applying Harris–Laplace salient point detector, which uses a Harris corner detector and subsequently the Laplacian for scale selection. Then a set of local descriptors, including gray OC-LBP, three best color OC-LBP, CS-LBP, SURF, HOG, SIFT and three best color SIFT, are extracted from 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 decreases the accuracy for object recognition.

5.3.3. Bag-of-Features modeling

After the step of feature extraction, each image is represented by a set of local descriptors. The number of local descriptors in each image varies because the number of the interest points (normally around thousands) changes from one image to another one. Thus, an efficient modeling method is required to transform

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

Fig. 9. Example images of the PASCAL VOC 2007 benchmark.

Fig. 10. Example images of the SIMPLIcity dataset.

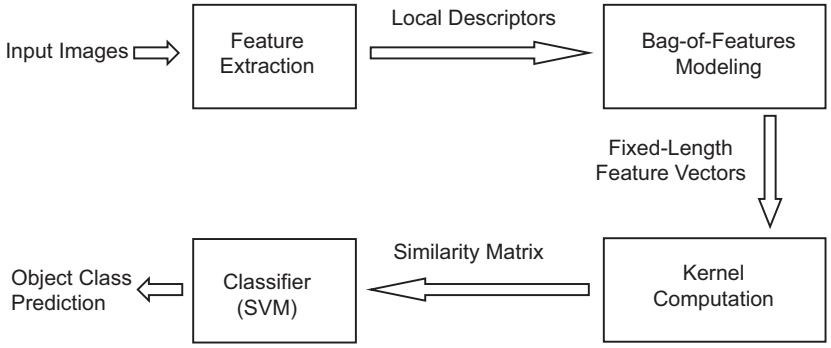

Fig. 11. Flowchart of our approach for object recognition.

this variable number of local descriptors into a more compact, informative and fixed-length representation for further classification.

We apply the popular Bag-of-Features (BoF) approach [5] because of its great success in object recognition tasks. As introduced in Section 1, the main idea of 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.

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

5.3.4. Classification

The support vector machine (SVM) algorithm is applied for object classification. Here the LibSVM implementation [45] is used. Once all the local descriptors are transformed to fixed-length feature vectors by the BoF method, the χ^2 distance is computed as Eq. (3) to measure the similarity between each pair of feature vectors. Then, the kernel function based on this distance is computed as Eq. (4) for the SVM training and prediction.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

Table 3
Object recognition results on the PASCAL VOC 2007 benchmark ("NOP-OC-LBP" is the abbreviation of "NOPPONENT-OC-LBP", "OP-SIFT" is the abbreviation of "OPPONENT-SIFT").

AP (%)	OC-LBP	Hue-OC-LBP	NOP-OC-LBP	RGB-OC-LBP	CS-LBP	HOG	SURF	SIFT	OP-SIFT	C-SIFT	RGB-SIFT
Airplane	62.2	64.3	64.2	61.9	59.2	52.1	39.7	56.0	59.9	58.7	57.8
Bicycle	38.6	35.4	39.1	42.0	44.8	26.9	45.9	44.9	43.8	38.9	44.6
Bird	25.9	32.9	34.8	32.1	27.4	25.0	26.7	28.2	27.7	32.1	22.5
Boat	56.4	56.0	60.8	59.5	53.0	40.6	21.0	45.7	49.1	51.8	46.6
Bottle	15.0	20.4	20.0	20.3	19.5	12.8	10.2	19.6	21.2	21.4	21.0
Bus	37.8	35.5	35.0	41.1	33.2	38.3	28.1	37.7	38.0	32.5	37.7
Car	62.6	60.5	61.4	65.1	63.1	58.1	52.5	55.0	57.4	53.2	56.1
Cat	38.9	39.3	39.7	42.9	40.2	27.5	24.3	36.5	37.7	34.1	37.3
Chair	39.0	40.5	41.3	39.3	38.7	43.8	33.3	44.5	42.4	45.9	43.5
Cow	20.6	21.5	14.6	24.9	18.3	19.8	20.8	25.9	17.0	16.6	27.8
Table	35.0	36.1	37.0	32.0	33.1	33.6	25.7	29.6	36.7	38.7	29.1
Dog	32.8	35.3	29.4	33.4	31.7	20.4	23.8	26.5	29.8	29.1	28.8
Horse	57.6	64.6	63.6	58.3	55.2	59.3	50.7	57.0	59.1	61.9	54.8
Motor	36.9	39.2	41.7	37.3	34.1	37.2	37.4	30.2	33.9	44.4	32.1
Person	74.1	77.2	75.5	74.7	73.0	66.2	70.8	73.1	74.5	76.6	72.7
Plant	21.3	22.7	26.7	20.1	17.5	10.4	13.8	11.5	19.9	27.1	11.5
Sheep	12.3	23.5	26.0	19.9	16.9	18.4	9.4	27.4	31.2	30.9	19.4
Sofa	25.8	27.8	27.5	25.0	19.0	26.3	19.3	23.6	22.9	23.2	24.6
Train	56.1	44.2	51.7	55.5	56.8	52.7	42.9	53.4	54.5	58.5	51.1
Monitor	25.6	29.2	27.9	31.8	31.7	32.3	25.7	33.7	35.0	27.3	35.6
Mean	**38.7**	**40.3**	**40.9**	**40.9**	**38.3**	**35.1**	**31.1**	**38.0**	**39.6**	**40.1**	**37.7**

For the SIMPLIcity dataset, each image is classified into the category with the maximum SVM output decision value. We tune the parameters of the classifier on the training set via 5-fold cross-validation, and obtain 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 the SVM classifier, and the AP (average precision) value 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 obtain the classification results on the test set.

5.3.5. Experimental results on PASCAL VOC 2007

The object recognition results on the PASCAL VOC 2007 benchmark are shown in Table 3. It can be seen that: (1) the proposed OC-LBP descriptor achieves the performance of 38.7% MAP, which is better than SURF and HOG, and comparable with CS-LBP and SIFT; (2) the best three color OC-LBP descriptors (Hue-OC-LBP, NOPPONENT-OC-LBP and RGB-OC-LBP) achieve 40.3%, 40.9% and 40.9% MAP, respectively, which outperform the intensity-based OC-LBP by about 2–3%, indicating that they truly benefit from additional color information and illumination invariance properties; (3) compared to the state-of-the-art color SIFT descriptors, the best three color OC-LBP descriptors achieve comparable or even better results.

After analyzing the detailed results in Table 3 by each object category, we could observe that the LBP-based descriptors generally perform better on the non-rigid object categories such as bird, cat, dog, horse, person, plant and sofa, while the SIFT-based descriptors are generally 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 further combine different color OC-LBP descriptors, as well as color OC-LBP and color SIFT by average late fusion to check if they can 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 OC-LBP and SIFT, proving that different color descriptors are not entirely redundant; (2) the color OC-LBP descriptors still achieve comparable or slightly better results than color SIFT after fusion; (3) the performance can be further improved (more than 4%) by fusing color OC-LBP and color SIFT, indicating that these two kinds of descriptors can provide complementary information to each other.

Table 4
Fusion results of color OC-LBP and color SIFT on the PASCAL VOC 2007 benchmark.

AP (%)	FUSION (3 color OC-LBP)	FUSION (3 color SIFT)	FUSION (3 color OC-LBP+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**

5.3.6. Experimental results on SIMPLIcity

The object recognition results on the SIMPLIcity dataset are shown in Tables 5 and 6. The similar observations to that on the PASCAL VOC benchmark can be noticed. The color OC-LBP descriptors outperform CS-LBP, SURF, HOG, SIFT, as well as the intensity-based OC-LBP, and achieve comparable results with the color SIFT descriptors. Further improvement (nearly 5%) can be obtained by fusing three color OC-LBP and three color SIFT

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

Table 5
Object recognition results on the SIMPLIcity dataset ("NOP-OC-LBP" is the abbreviation of "NOPPONENT-OC-LBP", "OP-SIFT" is the abbreviation of "OPPONENT-SIFT").

Accuracy (%)	OC-LBP	Hue-OC-LBP	NOP-OC-LBP	RGB-OC-LBP	CS-LBP	HOG	SURF	SIFT	OP-SIFT	C-SIFT	RGB-SIFT
People	70.0	84.0	80.0	78.0	70.0	58.0	72.0	76.0	76.0	84.0	74.0
Beach	74.0	82.0	86.0	76.0	82.0	68.0	76.0	82.0	88.0	86.0	82.0
Building	82.0	86.0	84.0	82.0	80.0	66.0	66.0	74.0	78.0	74.0	70.0
Bus	98.0	96.0	96.0	98.0	88.0	90.0	92.0	94.0	96.0	90.0	96.0
Dinosaur	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Elephant	74.0	70.0	72.0	72.0	80.0	70.0	78.0	88.0	84.0	74.0	94.0
Flower	82.0	94.0	88.0	86.0	88.0	58.0	70.0	92.0	96.0	86.0	88.0
Horse	98.0	98.0	98.0	96.0	96.0	92.0	82.0	96.0	98.0	100.0	94.0
Mountain	68.0	68.0	74.0	68.0	64.0	64.0	50.0	62.0	70.0	72.0	70.0
Food	88.0	92.0	100.0	96.0	80.0	72.0	78.0	86.0	88.0	94.0	90.0
Mean	**83.4**	**87.0**	**87.8**	**85.2**	**82.8**	**73.8**	**76.4**	**85.0**	**87.4**	**86.0**	**85.8**

Table 6
Fusion results of color OC-LBP and color SIFT on the SIMPLIcity dataset.

Accuracy (%)	FUSION (3 color OC-LBP)	FUSION (3 color SIFT)	FUSION (3 Color OC-LBP+3 Color SIFT)
People	86.0	86.0	86.0
Beach	86.0	88.0	86.0
Building	86.0	78.0	86.0
Bus	100.0	98.0	100.0
Dinosaur	100.0	100.0	100.0
Elephant	82.0	90.0	86.0
Flower	98.0	100.0	98.0
Horse	98.0	100.0	100.0
Mountain	78.0	76.0	82.0
Food	96.0	96.0	98.0
Mean	**91.0**	**91.2**	**92.2**

descriptors, since they provide complementary information to each other.

5.3.7. Performance comparison of the OC-LBP descriptor using SVM with different kernels

In order to compare and evaluate the effect of using SVM with different kernels for classification, we made supplementary comparisons in this section. The recognition approach remains the same as the one introduced in Section 5.3.1, and only the OC-LBP descriptor is used as feature. Different kernel comparisons are considered, including linear kernel vs. non-linear kernel, and comparison between different non-linear kernels (Chi-square (χ^2) vs. RBF). The results are presented in Tables 7 and 8.

From the results we can see that: (1) non-linear kernels perform much better than linear kernel, suggesting that the feature distributions of these datasets are quite complex, and linear kernel is not a good choice; (2) for different non-linear kernels, Chi-square (χ^2) performs better than RBF. The reason is that Chi-square (χ^2) is a histogram distance and Chi-square (χ^2) kernel outputs similarities between histogram-based features as the ones delivered by the Bag-of-Features (BoF) approach. Therefore, we choose Chi-square (χ^2) kernel together with the BoF approach for classification in our experiments.

Table 7
Performance comparison of the OC-LBP descriptor using SVM with different kernels on the PASCAL VOC 2007 benchmark.

AP (%)	Linear	RBF	Chi-square (χ^2)
Airplane	57.6	57.5	62.2
Bicycle	20.3	23.8	38.6
Bird	14.1	19.9	25.9
Boat	51.6	47.8	56.4
Bottle	10.0	11.3	15.0
Bus	22.7	33.9	37.8
Car	44.7	51.7	62.6
Cat	25.3	33.2	38.9
Chair	35.6	35.4	39.0
Cow	17.2	18.3	20.6
Table	17.7	24.9	35.0
Dog	21.9	28.9	32.8
Horse	44.0	53.4	57.6
Motor	23.4	24.2	36.9
Person	66.1	70.1	74.1
Plant	12.8	13.5	21.3
Sheep	7.9	7.4	12.3
Sofa	14.0	15.4	25.8
Train	36.8	48.9	56.1
Monitor	19.3	21.5	25.6
Mean	**28.2**	**32.1**	**38.7**

Table 8
Performance comparison of the OC-LBP descriptor using SVM with different kernels on the SIMPLIcity dataset.

Accuracy (%)	Linear	RBF	Chi-square (χ^2)
People	66.0	80.0	70.0
Beach	64.0	70.0	74.0
Building	60.0	74.0	82.0
Bus	96.0	98.0	98.0
Dinosaur	100.0	100.0	100.0
Elephant	78.0	70.0	74.0
Flower	74.0	84.0	82.0
Horse	86.0	96.0	98.0
Mountain	64.0	60.0	68.0
Food	86.0	82.0	88.0
Mean	**77.4**	**81.4**	**83.4**

5.4. Experiments on scene classification

We also evaluated the proposed descriptors in the application of scene classification. The dataset from Oliva and Torralba [42] is used, and denoted as OT scene dataset. It consists of 2688 color images from eight scene categories: coast (360 samples), forest (328 samples), mountain (374 samples), open country (410 samples), highway (260 samples), inside city (308 samples), tall building (356 samples) and street (292 samples). Fig. 12 shows some sample images of each category.

5.4.1. Experimental setup

For this scene classification problem, our approach is the same as the one used for object recognition, as described in Section 5.3.1, but with a different setting. Instead of detecting interest points in images using Harris–Laplace detectors, we apply the dense

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

Fig. 12. Example images of the OT scene dataset.

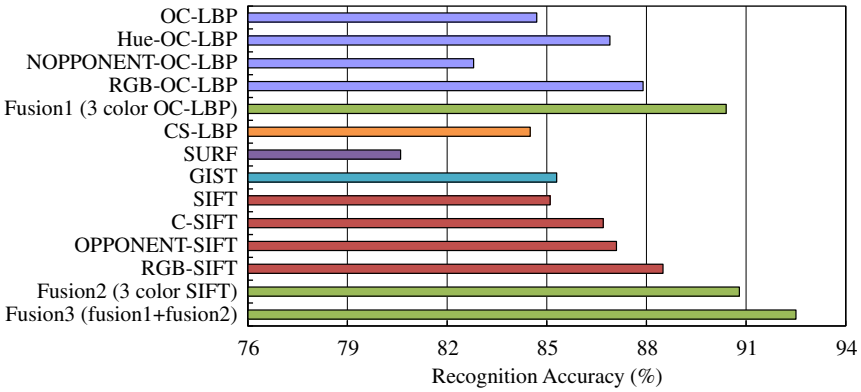

Fig. 13. Classification results on the OT scene dataset.

sampling strategy to locate keypoints for local descriptor computation. This is because for scene classification, we prefer to focus on the content of the whole image, rather than on "object" part only. Specifically, the sampling spacing is set to six 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 the corresponding Bag-of-Features (BoF) 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 five-fold cross-validation, and obtain the classification results on the test set.

5.4.2. Experimental results

The classification results on the OT scene dataset are shown in Fig. 13. It can be seen that the proposed OC-LBP descriptor performs better than the SURF descriptor, and achieves comparable results with the popular GIST, CS-LBP and SIFT. The proposed color OC-LBP descriptors further demonstrate their effectiveness as they display superior performances than all the intensity-based descriptors. They also show their ability of being complementary to the state-of-the-art color SIFT descriptors, since their fusion (fusion 3 in the figure) clearly improves the performance. It is worthy to notice that the NOPPONENT-OC-LBP descriptor does not perform well in this case, while its performance is quite good in the application of object recognition. The reason could be that the OT scene dataset contains more varieties of illumination changes than the object recognition datasets whereas the NOPPONENT-OC-LBP descriptor is only invariant to light intensity change. This also explains why RGB-OC-LBP and RGB-SIFT perform the best among the color descriptors, since they possess the strongest invariance properties (invariant to light color change and shift).

5.5. Computational cost comparison between descriptors

As we state in the introduction, a good local descriptor should be both discriminative and computationally efficient. The discriminative power of the proposed gray and color OC-LBP descriptors has been demonstrated by the previous experiments and applications, and they achieve comparable or even better performances than the state-of-the-art SIFT and color SIFT descriptors. In this section, we show the computational efficiency of the proposed descriptors also in comparison with the popular SIFT and color SIFT.

The comparisons are conducted on the four image datasets used in the previous experiments by utilizing a computer with Intel Core 2 Duo CPU @ 3.16 GHz and 3 GB RAM. We implement the gray and color OC-LBP descriptors by a mixture of C and Matlab,[1] and use the "ColorDescriptor" software [44] to compute

[1] The source code for computing the gray and color OC-LBP descriptors will be publicly available online at http://perso.ec-lyon.fr/liming.chen/ and http://perso.ec-lyon.fr/charles-edmond.bichot/.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

Table 9
Computational cost comparison between OC-LBP and SIFT descriptor.

Times (s)	Oxford (1000×700)	SIMPLIcity (384×256)	PASCAL (500×375)	OT Scene (256×256)
OC-LBP	0.2731	0.0615	0.1006	0.0420
Hue-OC-LBP	1.0651	0.1967	0.3165	0.1367
NOPPONENT-OC-LBP	0.8891	0.1808	0.2963	0.1173
RGB-OC-LBP	0.6763	0.1784	0.2883	0.1145
SIFT	1.0641	0.3276	0.4315	0.1613
C-SIFT	3.3041	0.9752	1.3109	0.4879
OPPONENT-SIFT	3.1964	0.9593	1.2966	0.4828
RGB-SIFT	3.1471	0.9551	1.2824	0.4773
Total (3 color OC-LBP)	2.6305	0.5559	0.9011	0.3685
Total (3 color SIFT)	9.6476	2.8896	3.8899	1.4480

SIFT and color SIFT. We record in Table 9 the average computation time required per image for each descriptor respectively.

It can be seen that the OC-LBP descriptor is about four times faster to compute than SIFT. When incorporating color information, the computations of color descriptors are about three times slower than the intensity-based descriptors, mainly because of the increasing channels. However, the color OC-LBP descriptors are still about four times faster than color SIFT. The proposed descriptors are thus better armed for large-scale problems.

6. Conclusions

In this paper, a new operator called the orthogonal combination of local binary patterns, denoted as OC-LBP, has first been proposed. It aims at reducing the dimensionality of the original LBP operator while keeping its discriminative power and computational efficiency.

We have also introduced several new local descriptors for image region description based on the proposed OC-LBP operator: the gray OC-LBP descriptor and six color OC-LBP descriptors, namely RGB-OC-LBP, NRGB-OC-LBP, OPPONENT-OC-LBP, NOPPONENT-OC-LBP, Hue-OC-LBP and TC-OC-LBP. The proposed descriptors incorporate color information to increase their discriminative power, and also to enhance their photometric invariance properties over various illumination changes.

The experiments in three different applications – image matching, object recognition and scene classification – show the effectiveness of the proposed descriptors. They outperform the popular SIFT, CS-LBP, HOG and SURF descriptors, and achieve comparable or even better performances than the state-of-the-art color SIFT descriptors. Meanwhile, they provide complementary information to SIFT, since further improvement can be obtained by fusing them.

Moreover, the proposed gray and color OC-LBP descriptors are about four times faster to compute than the SIFT and color SIFT descriptors respectively. Therefore, they are very promising for large-scale recognition problems.

In our future work, we are willing to extend the proposed OC-LBP to the description of video data. We also want to investigate visual object detection for improving visual object recognition.

Conflict of interest statement

None declared.

Acknowledgments

This work was supported in part by the French research agency ANR through the VideoSense project under the grant 2009 CORD 026 02.

References

[1] M.J. Swain, D.H. Ballard, Color indexing, International Journal of Computer Vision (IJCV) 7 (1) (1991) 11–32.
[2] M.A. Stricker, M. Orengo, Similarity of color images, in: Proceedings of the Storage and Retrieval for Image and Video Databases (SPIE), 1995, pp. 381–392.
[3] D.K. Park, Y.S. Jeon, C.S. Won, Efficient use of local edge histogram descriptor, in: Proceedings of the ACM Multimedia Workshops, 2000, pp. 51–54.
[4] M. Tuceryan, A.K. Jain, Texture analysis, in: C. Chen, L. Pau, P. Wang (Eds.), Handbook of Pattern Recognition and Computer Vision, second ed.,World Scientific Publishing Co., River Edge, NJ, USA, 1998, pp. 207–248.
[5] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in: Proceedings of the Workshop on Statistical Learning in Computer Vision, ECCV, 2004, pp. 1–22.
[6] M. Everingham, L.J.V. Gool, C.K.I. Williams, J.M. Winn, A. Zisserman, The Pascal visual object classes (VOC) challenge, International Journal of Computer Vision (IJCV) 88 (2) (2010) 303–338.
[7] S. Nowak, M.J. Huiskes, New strategies for image annotation: overview of the photo annotation task at imageclef 2010, in: CLEF Workshop Notebook Paper, 2010.
[8] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision (IJCV) 60 (2) (2004) 91–110.
[9] Y. Ke, R. Sukthankar, Pca-sift: a more distinctive representation for local image descriptors, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004, pp. 506–513.
[10] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 27 (10) (2005) 1615–1630.
[11] H. Bay, A. Ess, T. Tuytelaars, L.J.V. Gool, Surf: speeded-up robust features, Computer Vision and Image Understanding (CVIU) 110 (3) (2008) 346–359.
[12] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.
[13] E. Tola, V. Lepetit, P. Fua, Daisy: an efficient dense descriptor applied to wide-baseline stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 32 (5) (2010) 815–830.
[14] B. Li, R. Xiao, Z. Li, R. Cai, B.-L. Lu, L. Zhang, Rank-sift: learning to rank repeatable local interest points, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1737–1744.
[15] M. Calonder, V. Lepetit, M. Özuysal, T. Trzcinski, C. Strecha, P. Fua, Brief: computing a local binary descriptor very fast, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 34 (7) (2012) 1281–1298.
[16] E. Rublee, V. Rabaud, K. Konolige, G.R. Bradski, Orb: an efficient alternative to sift or surf, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2011, pp. 2564–2571.
[17] Z. Wang, B. Fan, F. Wu, Local intensity order pattern for feature description, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2011, pp. 603–610.
[18] B. Fan, F. Wu, Z. Hu, Aggregating gradient distributions into intensity orders: a novel local image descriptor, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 2377–2384.
[19] J. Zhang, M. Marszalek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, International Journal of Computer Vision (IJCV) 73 (2) (2007) 213–238.
[20] J. Li, N.M. Allinson, A comprehensive review of current local features for computer vision, Neurocomputing 71 (10–12) (2008) 1771–1787.
[21] A.E. Abdel-Hakim, A.A. Farag, Csift: a sift descriptor with color invariant characteristics, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 1978–1983.
[22] A. Bosch, A. Zisserman, X. Muñoz, Scene classification using a hybrid generative/discriminative approach, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 30 (4) (2008) 712–727.
[23] J. van de Weijer, T. Gevers, A.D. Bagdanov, Boosting color saliency in image feature detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 28 (1) (2006) 150–156.
[24] G.J. Burghouts, J.-M. Geusebroek, Performance evaluation of local color invariants, Computer Vision and Image Understanding (CVIU) 113 (1) (2009) 48–62.
[25] K.E.A. van de Sande, T. Gevers, C.G.M. Snoek, Evaluating color descriptors for object and scene recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 32 (9) (2010) 1582–1596.
[26] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1) (1996) 51–59.
[27] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24 (7) (2002) 971–987.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003

[28] T. Mäenpää, T. Ojala, M. Pietikäinen, M. Soriano, Robust texture classification by subsets of local binary patterns, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2000, pp. 3947–3950.
[29] T. Mäenpää, M. Pietikäinen, T. Ojala, Texture classification by multi-predicate local binary pattern operators, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2000, pp. 3951–3954.
[30] T. Ojala, M. Pietikäinen, Unsupervised texture segmentation using feature distributions, Pattern Recognition 32 (3) (1999) 477–486.
[31] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: Proceedings of the European Conference on Computer Vision (ECCV), 2004, pp. 469–481.
[32] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 28 (12) (2006) 2037–2041.
[33] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image and Vision Computing (IVC) 27 (6) (2009) 803–816.
[34] G. Zhao, M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI 29 (6) (2007) 915–928.
[35] M. Heikkilä, M. Pietikäinen, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition 42 (3) (2009) 425–436.
[36] C. Zhu, C.-E. Bichot, L. Chen, Multi-scale color local binary patterns for visual object classes recognition, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2010, pp. 3065–3068.
[37] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 41 (4) (2011) 1–17.
[38] T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, S. Huovinen, Outex—new framework for empirical evaluation of texture analysis algorithms, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2002, pp. 701–706.
[39] U.o.O. Visual Geometry Group, Comparison of Region Descriptors ⟨http://www.robots.ox.ac.uk/~vgg/research/affine/desc_evaluation.html⟩.
[40] M. Everingham, L.V. Gool, C. Williams, J. Winn, A. Zisserman, The Pascal Visual Object Classes Challenge 2007 (voc2007) Results ⟨http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html⟩.
[41] J.Z. Wang, J. Li, G. Wiederhold, Simplicity: semantics-sensitive integrated matching for picture libraries, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 23 (9) (2001) 947–963.
[42] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, International Journal of Computer Vision (IJCV) 42 (3) (2001) 145–175.
[43] K. Mikolajczyk, C. Schmid, An affine invariant interest point detector, in: Proceedings of the European Conference on Computer Vision (ECCV), 2002, pp. 128–142.
[44] U.o.A. Koen van de Sande, Color Descriptor Software ⟨http://www.colordescriptors.com⟩.
[45] C. Chung Chang, C.-J. Lin, Libsvm: A Library for Support Vector Machines ⟨http://www.csie.ntu.edu.tw/cjlin/libsvm⟩ 2001.

Chao Zhu received his bachelor diploma of automation in 2005 from the University of Electronic and Technology of Xi'an, China, his Master Degree of System Engineering in 2008 from Xi'an Jiaotong University, China, and his PhD degree of computer science in 2012 from Ecole Centrale de Lyon, France. Currently, he works as a post-doc at Ecole Centrale de Lyon. His research interests include object recognition, image classification, feature extraction and image processing.

Charles-Edmond Bichot received in 2004 both an aeronautical engineering degree and an MSc degree in computer science from the Ecole Nationale de l'Aviation Civile and the Université Toulouse III, France. He obtained in 2007 a PhD in computer science from the Institute National Polytechnique de Toulouse (INPT) which was awarded by the INPT PhD Thesis prize and by the Aerospace-Valley prize in 2008. After 1 year as an INRIA postdoctoral research assistant at the LaBRI laboratory in Bordeaux, he joined as associate professor at the Ecole Centrale de Lyon and the LIRIS laboratory, France. His research interests include object recognition, image segmentation, graph partitioning, combinatorial optimization and UAV.

Liming Chen was awarded a joint BSc degree in mathematics and computer science from the University of Nantes in 1984. He obtained a master degree in 1986 and a PhD in computer science from the University of Paris 6 in 1989. He first served as associate professor at the Université de Technologie de Compiègne, then joined Ecole Centrale de Lyon as Professor in 1998, where he leads an advanced research team on multimedia computing and pattern recognition. From 2001 to 2003, he also served as Chief Scientific Officer in a Paris-based company, Avivias, specialized in media asset management. In 2005, he served as scientific multimedia expert in France Telecom R&D China. He has been head of the department of mathematics and computer science from 2007. Prof. Liming Chen has taken out three patents, authored more than 100 publications and acted as chairman, PC member and reviewer in a number of high profile journal and conferences since 1995. He has been a (co)-principal investigator on a number of research grants from EU FP program, French research funding bodies and local government departments. He has directed more than 15 PhD theses. His current research spans from 2D/3D face analysis and recognition, image and video analysis and categorization, to affect analysis both in image, audio and video.

Please cite this article as: C. Zhu, et al., Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.01.003