Chapter 10
LBP and Color Descriptors for Image Classification

Sugata Banerji, Abhishek Verma, and Chengjun Liu

Abstract. Four novel color Local Binary Pattern (LBP) descriptors are presented in this chapter for scene image and image texture classification with applications to image search and retrieval. Specifically, the first color LBP descriptor, the oRGB-LBP descriptor, is derived by concatenating the LBP features of the component images in an opponent color space — the oRGB color space. The other three color LBP descriptors are obtained by the integration of the oRGB-LBP descriptor with some additional image features: the Color LBP Fusion (CLF) descriptor is constructed by integrating the RGB-LBP, the YCbCr-LBP, the HSV-LBP, the rgb-LBP, as well as the oRGB-LBP descriptor; the Color Grayscale LBP Fusion (CGLF) descriptor is derived by integrating the grayscale-LBP descriptor and the CLF descriptor; and the CGLF+PHOG descriptor is obtained by integrating the Pyramid of Histograms of Orientation Gradients (PHOG) and the CGLF descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) and image classification is based on the nearest neighbor classification rule (EFM-NN). The proposed image descriptors and the feature extraction and classification methods are evaluated using three databases: the MIT scene database, the KTH-TIPS2-b database, and the KTH-TIPS materials database. The experimental results show that (i) the proposed oRGB-LBP descriptor improves image classification performance upon other color LBP descriptors, and (ii) the CLF, the CGLF, and the CGLF+PHOG descriptors further improve upon the oRGB-LBP descriptor for scene image and image texture classification.

10.1 Introduction

Color features have been shown to achieve higher success rate than grayscale features in image search and retrieval due to the fact that color features contain

Sugata Banerji · Abhishek Verma · Chengjun Liu
New Jersey Institute of Technology, Newark, NJ 07102, USA
e-mail: {sb256,av56,chengjun.liu}@njit.edu

C. Liu and V.K. Mago: Cross Disciplinary Biometric Systems, ISRL 37, pp. 205–225.
springerlink.com

significantly larger amount of discriminative information [32, 48, 41, 45]. Color based image search can be very useful in the identification of object and natural scene categories [45, 46, 2]. Color features can be derived from various color spaces and they exhibit different properties. Two necessary properties for color feature detectors are that they need to be stable under changing viewing conditions, such as changes in illumination, shading, highlights, and they should have a high discriminative power. Global color features such as the color histogram and local invariant features provide varying degrees of success against image variations such as rotation, viewpoint and lighting changes, clutter and occlusions [7, 43].

In recent years, the recognition and classification of textures using the Local Binary Pattern (LBP) features has been shown to be promising [36, 37, 50, 9, 10]. Color features when combined with the intensity based texture descriptors are able to outperform many alternatives. In this chapter, a variable mask size is employed to generate a multi-scale LBP feature vector that is more robust to changes in scale and orientation. Furthermore, the multi-scale LBP descriptor is extended to different color spaces including the recently proposed oRGB color space [6] and a new multi-scale oRGB-LBP feature representation is proposed. The new oRGB-LBP descriptor is then integrated with other color LBP features to produce the novel multi-scale Color LBP Fusion (CLF) and the multi-scale Color Grayscale LBP Fusion (CGLF) descriptors. The CGLF is further combined with the Pyramid of Histograms of Orientation Gradients (PHOG) to obtain the novel CGLF+PHOG descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) [28, 30] and image classification is based on the nearest neighbor classification rule (EFM-NN) [13]. The effectiveness of the proposed descriptors and the EFM-NN classification method is shown using three datasets: the MIT scene database, the KTH-TIPS2-b and the KTH-TIPS materials databases.

10.2 Related Work

The use of color as a means to image retrieval [32, 25, 41] and object and scene search [45] has gained popularity recently. Color features can capture discriminative information by means of the color invariants, color histogram, color texture, etc. The early methods for object and scene classification were mainly based on the global descriptors such as the color and texture histograms [35, 39, 40]. One representative method is the color indexing system designed by Swain and Ballard, which uses the color histogram for image inquiry from a large image database [44]. These early methods are sensitive to viewpoint and lighting changes, clutter and occlusions. For this reason, global methods were gradually replaced by the part-based methods, which became one of the popular techniques in the object recognition community. Part-based models combine appearance descriptors from local features along with their spatial relationship. Harris interest point detector, for example, was used for local feature extraction, but such features are only invariant to translation [1, 47]. Afterwards, local features with greater invariance were developed, which

were found to be robust against scale changes [12] and affine deformations [20]. Learning and inference for spatial relations poses a challenging problem in terms of its complexity and computational cost. Whereas, the orderless bag-of-words methods [12, 21, 18] are simpler and computationally efficient, though they are not able to represent the geometric structure of the object or to distinguish between foreground and background features. For these reasons, the bag-of-words methods are not robust to clutter. One way to overcome this drawback is to design kernels that can yield high discriminative power in presence of noise and clutter [16].

More recent work on color based image classification appears in [32, 48, 45, 26] that propose several new color spaces and methods for face, object and scene classification. The HSV color space is used for scene category recognition in [5], and the evaluation of local color invariant descriptors is performed in [7]. Fusion of color models, color region detection and color edge detection has been investigated for representation of color images [43]. Some important contributions of color, texture, and shape abstraction for image retrieval have been discussed in Datta et al. [11].

Many recent techniques for the description of images have considered local features, and one representative local image descriptor is the Scale-Invariant Feature Transform (SIFT) [33]. The SIFT descriptor, which encodes the distribution of Gaussian gradients within an image region, can efficiently represent the spatial intensity pattern and is robust to small deformations and localization errors. Currently, several modifications to the SIFT features have been proposed, such as the Gradient Location and Orientation Histogram (GLOH) [34], and the Speeded-Up Robust Features (SURF) [3]. These region-based descriptors have achieved a high degree of invariance to the overall illumination conditions for planar surfaces. Although designed to retrieve identical object patches, SIFT-like features turn out to be quite successful in the bag-of-words approaches for general scene and object classification [5].

Lately, several methods based on LBP features have been proposed for image representation and classification [50, 10]. In a 3×3 neighborhood of an image, the basic LBP operator assigns a binary label 0 or 1 to each surrounding pixel by thresholding at the gray value of the central pixel and replacing its value with a decimal number converted from the 8-bit binary number. Extraction of the LBP features is computationally efficient and with the use of multi-scale filters their invariance to scaling and rotation can be achieved [50]. Fusion of different LBP features has been shown to achieve a good image retrieval success rate [2, 10, 49]. Local image descriptors have also been shown to perform well for texture based image retrieval [2, 9, 49].

The Pyramid of Histograms of Orientation Gradients (PHOG) descriptor [4] is able to represent an image by its local shape and the spatial layout of the shape. The local shape is captured by the distribution over edge orientations within a region, and the spatial layout by tiling the image into regions at multiple resolutions. The distance between two PHOG image descriptors then reflects the extent to which the images contain similar shapes and correspond in their spatial layout. Figure 10.1 shows how the PHOG descriptor is formed by the concatenation of the gradient histograms over different resolutions from a scene image.

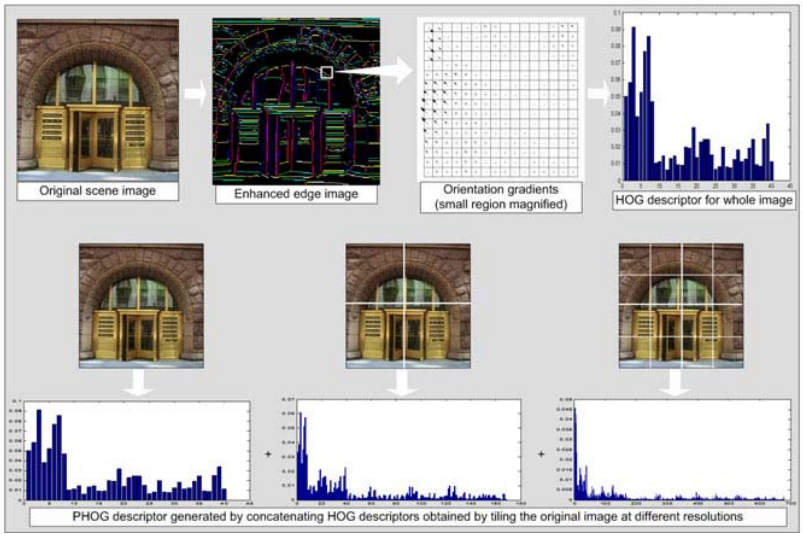

Fig. 10.1 The Pyramid Histograms of Orientation Gradients (PHOG) descriptor.

Efficient image retrieval requires a robust feature extraction method that has the ability to learn meaningful low-dimensional patterns in spaces of very high dimensionality [22, 27, 31]. Low-dimensional representation is also important when one considers the intrinsic computational aspect. Principal Component Analysis (PCA) has been widely used to perform dimensionality reduction for image indexing and retrieval [28, 24]. The EFM feature extraction method has achieved good success for the task of image based representation and retrieval [30, 29, 23]. Efficient image retrieval also requires an image classification method. Recently, the Support Vector Machine (SVM) classifier has been applied for multiple category recognition [49]. However, the SVM classifier suffers from the drawback of being computationally expensive on large scale image classification tasks. An alternative method is to apply the EFM feature extraction method and the Nearest Neighbor classification rule (EFM-NN) for image classification, namely the EFM-NN classifier.

10.3 Color Spaces and the New Color LBP Descriptors

We review in this section five color spaces and then define four new color LBP descriptors: the new oRGB-LBP descriptor, the Color LBP Fusion (CLF) descriptor, the Color Grayscale LBP Fusion (CGLF) descriptor, and the CGLF+PHOG descriptor. In comparison, the conventional LBP descriptor forms the intensity-based (grayscale) LBP descriptor.

A color image contains three component images, and each pixel of a color image is specified in a color space, which serves as a color coordinate system. The commonly used color space is the RGB color space. Other color spaces are usually calculated from the RGB color space by means of either linear or nonlinear transformations. To reduce the sensitivity of the RGB images to luminance, surface orientation, and other photographic conditions, the rgb color space is defined by normalizing the R, G, and B components:

$$\begin{aligned} r &= R/(R+G+B) \\ g &= G/(R+G+B) \\ b &= B/(R+G+B) \end{aligned} \tag{10.1}$$

Due to the normalization r and g are scale-invariant and thereby invariant to light intensity changes, shadows and shading [14].

The HSV color space is motivated by the human vision system because humans describe color by means of hue, saturation, and brightness. Hue and saturation define chrominance, while intensity or value specifies luminance [15]. The HSV color space is defined as follows [42]:

$$\begin{aligned} &Let \quad \begin{cases} MAX = max(R,G,B) \\ MIN = min(R,G,B) \\ \delta = MAX - MIN \end{cases} \\ &V = MAX \\ &S = \begin{cases} \frac{\delta}{MAX} & \text{if } MAX \neq 0 \\ 0 & \text{if } MAX = 0 \end{cases} \\ &H = \begin{cases} 60(\frac{G-B}{\delta}) & \text{if } MAX = R \\ 60(\frac{B-R}{\delta}+2) & \text{if } MAX = G \\ 60(\frac{R-G}{\delta}+4) & \text{if } MAX = B \\ not\ defined & \text{if } MAX = 0 \end{cases} \end{aligned} \tag{10.2}$$

The YCbCr color space is developed for digital video standard and television transmissions. In YCbCr, the RGB components are separated into luminance, chrominance blue, and chrominance red:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{10.3}$$

where the R,G,B values are scaled to $[0,1]$.

The oRGB color space [6] has three channels L, $C1$ and $C2$. The primaries of this model are based on the three fundamental psychological opponent axes: white-black, red-green, and yellow-blue. The color information is contained in $C1$ and $C2$. The values of $C1$ are within $[-1,1]$ and the values of $C2$ are within

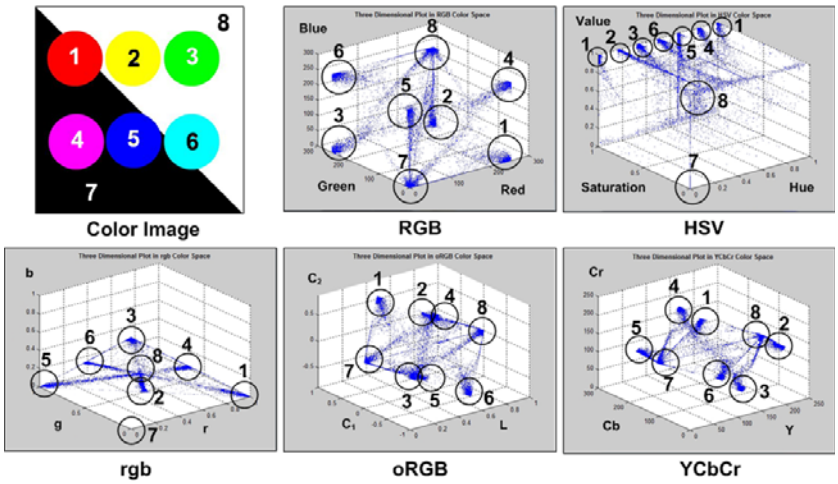

Fig. 10.2 Visualizing eight different colors in various color spaces. The upper left figure is the image with eight colors numbered from 1 to 8. The remaining five figures display the eight colors in the RGB space, the HSV space, the rgb space, the oRGB space, and the YCbCr space, respectively.

$[-0.8660, 0.8660]$. The L channel contains the luminance information and its values are within $[0, 1]$:

$$\begin{bmatrix} L \\ C1 \\ C2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{10.4}$$

Figure 10.2 shows eight different colors in various color spaces. Figure 10.3 shows the color component images in the five color spaces: RGB, HSV, rgb, oRGB, and YCbCr.

The LBP descriptor [36, 37] assigns an intensity value to each pixel of an image based on the intensity values of its eight neighboring pixels. Choosing multiple neighborhoods of different distances from the target pixel and orientations for each pixel has been shown to achieve partial invariance to scaling and rotation [50]. Using the multi-scale LBP operator shown in Figure 10.4, three LBP images are generated from the three neighborhoods. The normalized histograms from the LBP images are used as feature vectors and they are independent of the image size. The fused histograms of multi-scale LBP images give a feature vector that is partially invariant to image translation, scaling, and rotation.

The grayscale-LBP descriptor is defined as the LBP descriptor applied to the grayscale image. A color LBP descriptor in a given color space is derived by individually computing the LBP descriptor on each of the three component images in the specific color space. This produces a 2304 dimensional descriptor that is formed

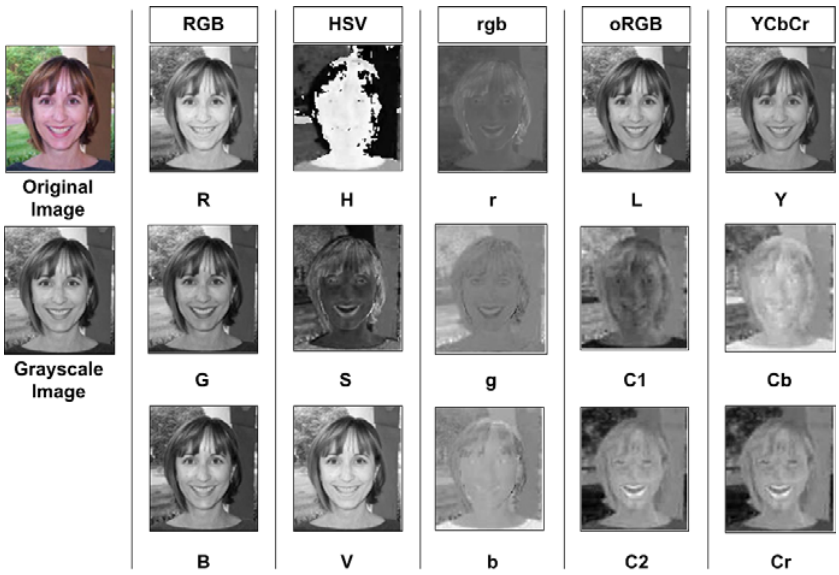

Fig. 10.3 Color component images in the five color spaces: RGB, HSV, rgb, oRGB, and YCbCr. The color image is from the Caltech 256 dataset, whose grayscale image is displayed as well.

from concatenating the 768 dimensional vectors from the three channels. As a result, four color LBP descriptors are defined as follows: the RGB-LBP descriptor, the YCbCr-LBP descriptor, the HSV-LBP descriptor, and the rgb-LBP descriptor.

Four new color LBP descriptors are defined in the oRGB color space and by the fusion of descriptors in different color spaces, respectively. In particular, the oRGB-LBP descriptor is constructed by concatenating the LBP descriptors of the three component images in the oRGB color space. The Color LBP Fusion (CLF) descriptor is formed by fusing the RGB-LBP, the YCbCr-LBP, the HSV-LBP, the oRGB-LBP, and the rgb-LBP descriptors. The Color Grayscale LBP Fusion (CGLF) descriptor is obtained by fusing the CLF descriptor and the grayscale-LBP descriptor. And finally the CGLF+PHOG descriptor is formed by combining the CGLF with the PHOG.

10.4 The EFM-NN Classifier

Image classification using the new color LBP descriptors introduced in the preceding section is implemented using the Enhanced Fisher Model (EFM) feature extraction method [30, 28] and the Nearest Neighbor classification rule (EFM-NN),

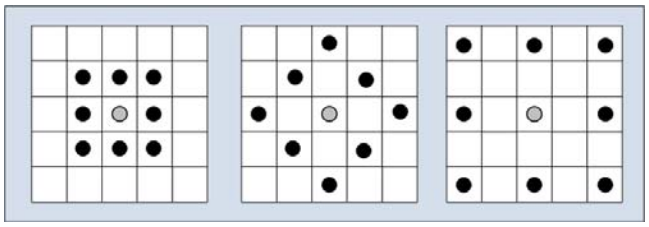

Fig. 10.4 The different neighborhoods of the multi-scale LBP operators.

i.e., the EFM-NN classifier. Let $\mathscr{X} \in \mathbb{R}^N$ be a random vector whose covariance matrix is $\Sigma_{\mathscr{X}}$:

$$\Sigma_{\mathscr{X}} = \mathscr{E}\{[\mathscr{X} - \mathscr{E}(\mathscr{X})][\mathscr{X} - \mathscr{E}(\mathscr{X})]^t\} \tag{10.5}$$

where $\mathscr{E}(\cdot)$ is the expectation operator and t denotes the transpose operation. The eigenvectors of the covariance matrix $\Sigma_{\mathscr{X}}$ can be derived by means of PCA:

$$\Sigma_{\mathscr{X}} = \Phi \Lambda \Phi^t \tag{10.6}$$

where $\Phi = [\phi_1 \phi_2 \ldots \phi_N]$ is an orthogonal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ a diagonal eigenvalue matrix with the diagonal elements in decreasing order. An important application of PCA is dimensionality reduction:

$$\mathscr{Y} = P^t \mathscr{X} \tag{10.7}$$

where $P = [\phi_1 \phi_2 \ldots \phi_K]$, and $K < N$. $\mathscr{Y} \in \mathbb{R}^K$ thus is composed of the most significant principal components. PCA, which is derived based on an optimal representation criterion, usually does not lead to good image classification performance. To improve upon PCA, the Fisher Linear Discriminant (FLD) analysis [13] is introduced to extract the most discriminating features.

The FLD method optimizes a criterion defined on the within-class and between-class scatter matrices, S_w and S_b [13]:

$$S_w = \sum_{i=1}^{L} P(\omega_i) \mathscr{E}\{(\mathscr{Y} - M_i)(\mathscr{Y} - M_i)^t | \omega_i\} \tag{10.8}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t \tag{10.9}$$

where $P(\omega_i)$ is *a priori* probability, ω_i represent the classes, and M_i and M are the means of the classes and the grand mean, respectively. The criterion the FLD method optimizes is $J_1 = tr(S_w^{-1} S_b)$, which is maximized when Ψ contains the eigenvectors of the matrix $S_w^{-1} S_b$ [13]:

$$S_w^{-1} S_b \Psi = \Psi \Delta \tag{10.10}$$

where Ψ, Δ are the eigenvector and eigenvalue matrices of $S_w^{-1}S_b$, respectively. The FLD discriminating features are defined by projecting the pattern vector $\mathscr{Y}$ onto the eigenvectors in Ψ:

$$\mathscr{Z} = \Psi^t \mathscr{Y} \tag{10.11}$$

$\mathscr{Z}$ thus is more effective than the feature vector $\mathscr{Y}$ derived by PCA for image classification.

The FLD method, however, often leads to overfitting when implemented in an inappropriate PCA space. To improve the generalization performance of the FLD method, a proper balance between two criteria should be maintained: the energy criterion for adequate image representation and the magnitude criterion for eliminating the small-valued trailing eigenvalues of the within-class scatter matrix [28]. A new method, the Enhanced Fisher Model (EFM), is capable of improving the generalization performance of the FLD method [28]. Specifically, the EFM method improves the generalization capability of the FLD method by decomposing the FLD procedure into a simultaneous diagonalization of the within-class and between-class scatter matrices [28]. The simultaneous diagonalization is stepwise equivalent to two operations [13]: whitening the within-class scatter matrix and diagonalizing the between-class scatter matrix using the transformed data. The stepwise operation shows that during whitening the eigenvalues of the within-class scatter matrix appear in the denominator. Since the small (trailing) eigenvalues tend to capture noise [28], they cause the whitening step to fit for misleading variations, which leads to poor generalization performance. To achieve enhanced performance, the EFM method preserves a proper balance between the need that the selected eigenvalues account for most of the spectral energy of the raw data (for representational adequacy), and the requirement that the eigenvalues of the within-class scatter matrix are not too small (for better generalization performance) [28].

After feature extraction, image classification is implemented using the nearest neighbor classification rule. Figure 10.5 shows the multiple features fusion methodology that applies multiple color spaces (oRGB, YCbCr, RGB, HSV, rgb), multiscale LBP, various color LBP descriptors (oRGB-LBP, CLF, CGLF, CGLF+PHOG), PCA for dimensionality reduction, EFM for feature extraction, and the EFM-NN classifier for image classification.

10.5 Experiments

10.5.1 Datasets and Experimental Methodology

The following three publicly accessible datasets are used to evaluate the proposed color LBP descriptors and the EFM-NN classification method: the MIT Scene dataset [38], the KTH-TIPS (Textures under varying Illumination, Pose and Scale) [17] and KTH-TIPS2-b datasets [8]. The MIT scene dataset [38] has 2,688 images classified as eight categories: 360 coast, 328 forest, 374 mountain, 410 open

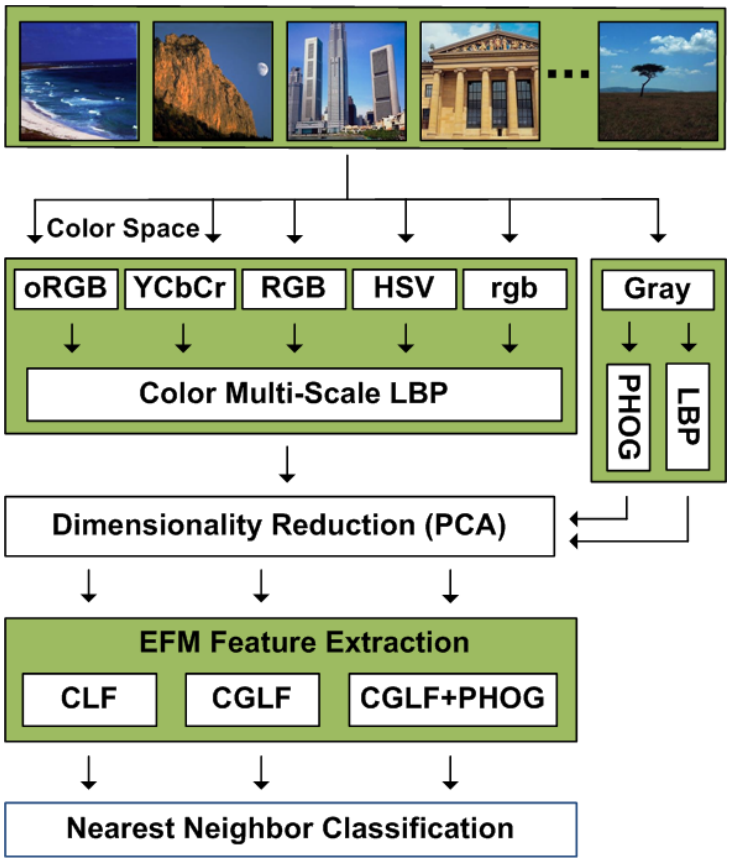

Fig. 10.5 Multiple feature fusion methodology that applies multiple color spaces (oRGB, YCbCr, RGB, HSV, rgb), multi-scale LBP, various color LBP descriptors (oRGB-LBP, CLF, CGLF, CGLF+PHOG), PCA for dimensionality reduction, EFM for feature extraction, and the EFM-NN classifier for image classification.

country, 260 highway, 308 inside of cities, 356 tall buildings, and 292 streets. All of the images are in color, in JPEG format, and the average size of each image is 256x256 pixels. There is a large variation in light, pose and angles, along with a high intra-class variation. The sources of the images vary (from commercial databases, websites, and digital cameras) [38]. See Figure 10.6(a) for some sample images from this dataset. The KTH-TIPS dataset [17, 8, 19] consists of 10 classes of textures with 81 images per class. All the images are in color, PNG format and the maximum image size is 200x200 pixels. All ten textures have been photographed at nine scales and nine illumination conditions for each scale. Some of the classes have a very similar visual appearance, like cotton and linen, and brown bread and sponge

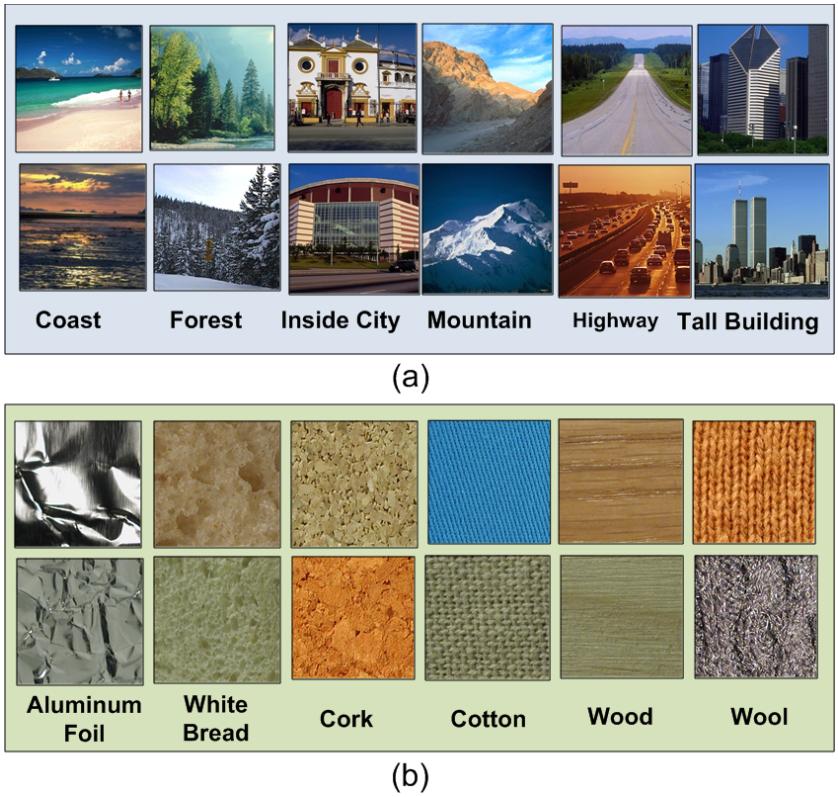

Fig. 10.6 Example images from (a) the MIT scene dataset and (b) the KTH-TIPS2-b materials dataset.

which makes this dataset moderately challenging. The KTH-TIPS2-b dataset [8] is a more challenging extension of the KTH-TIPS dataset with 11 classes of materials and 4 samples for each material. Each of these samples has 108 images with 432 images per class and a total of 4752 images. Some of the images in the classes like wool and cotton are from differently colored samples leading to very high intra-class variation between samples, while some samples from different classes like cork and cracker have the same color and general appearance lowering the inter-class variation. See Figure 10.6(b) for some sample images from this dataset.

The classification task is to assign each test image to one of a number of categories. The performance is measured using a confusion matrix, and the overall performance rates are measured by the average value of the diagonal entries of the confusion matrix. For the KTH-TIPS2-b dataset five random sets of 200 training images per class and 100 testing images per class are used. For the KTH-TIPS dataset five random sets of 40 training images per class and 41 test images per class are

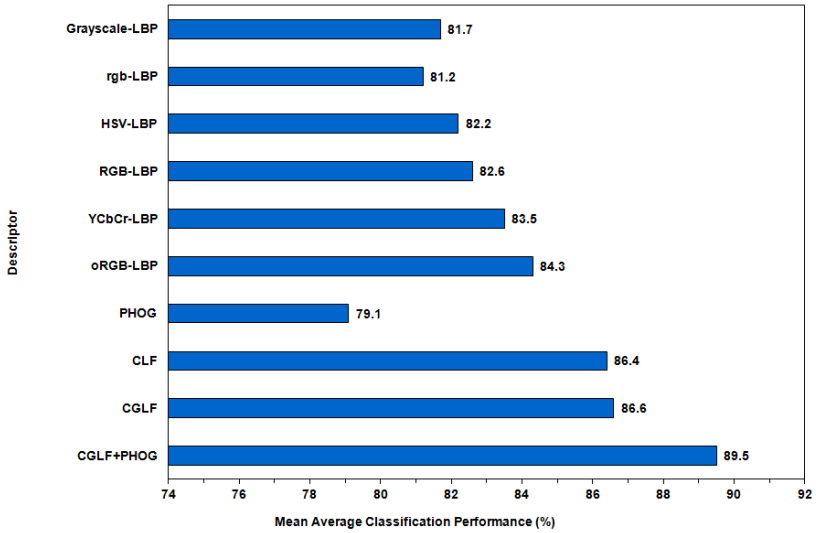

Fig. 10.7 The mean average classification performance of the ten descriptors using the EFM-NN classifier on the MIT scene dataset: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the PHOG, the CLF, the CGLF, and the CGLF+PHOG descriptors.

selected (same numbers as used in [10, 49, 19]). For the MIT scene dataset five image sets are randomly selected. Each set consists of 2000 images for training (250 images per class) and the rest 688 images for testing. Within each set there is no overlap in the images selected for training and testing. The classification scheme on the datasets compares the overall and category wise performance of ten different descriptors: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the CLF, the CGLF, the PHOG and the CGLF+PHOG descriptors (the final two evaluated on the scene dataset). Classification is implemented using the EFM-nearest neighbor (EFM-NN) classifier.

10.5.2 Evaluation of Novel Color Descriptors and EFM-Nearest Neighbor Classifier on the MIT Scene Dataset

The first set of experiments assesses the overall classification performance of the ten descriptors. Note that for each category five-fold cross validation is implemented for each descriptor using the EFM-nearest neighbor classifier to derive the average classification performance. As a result, each descriptor yields 8 average classification

Table 10.1 Category Wise Descriptor Performance (%) Split-out with the EFM-NN Classifier on the MIT Scene Dataset. Note That the Categories are Sorted on the CGLF+PHOG Results

Category	**CGLF+ PHOG**	**CGLF**	**CLF**	**oRGB LBP**	**YCbCr LBP**	**RGB LBP**	**HSV LBP**	**rgb LBP**	**Gray LBP**	**PHOG**
Highway	**97**	90	93	90	87	90	90	90	93	90
Forest	96	**97**	**97**	**97**	**97**	95	94	94	94	94
Coast	**91**	88	87	85	88	83	81	82	86	84
Street	**90**	**90**	86	83	83	82	84	82	81	86
Mountain	**90**	85	84	80	81	80	80	76	77	75
Tall Building	**90**	86	86	86	83	84	82	80	79	70
Inside City	86	**87**	**87**	86	83	81	80	79	83	79
Open Country	**76**	71	71	68	66	65	66	68	61	56
Mean	**89.5**	**86.6**	**86.4**	**84.2**	**83.5**	**82.6**	**82.2**	**81.2**	**81.7**	**79.1**

rates corresponding to the 8 image categories. The mean value of these 8 average classification rates is defined as the mean average classification performance for the descriptor. Figure 10.7 shows the mean average classification performance of various descriptors. The best classification rate is 89.5% from the CGLF+PHOG, which is good performance for a dataset of this size and complexity. The oRGB-LBP achieves the classification rate of 84.3%. It outperforms the other color LBP descriptors. It is noted that fusion of the color LBP descriptors (CLF) improves upon the grayscale-LBP by a significant 4.7% margin. The grayscale-LBP descriptor improves the fusion (CGLF) result slightly upon the CLF descriptor.

The second set of experiments assesses the ten descriptors using the EFM-nearest neighbor classifier on individual image categories. From Table 10.1 it can be seen that the top six categories achieve a success rate of over 90%. The Forest category achieves a success rate of over 90% across all ten descriptors. Individual color LBP features improve upon the grayscale-LBP on most of the categories. The CLF results on each of the eight categories show significant improvement upon the grayscale-LBP and the CGLF slightly improves upon the CLF. Integration of PHOG with the CGLF to obtain the CGLF+PHOG highly benefits most categories and in particular there is a significant increase in the classification performance upon the CGLF results for the Highway, Mountain and Open Country categories where the increment is in the range of 5% to 7%.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Figure 10.8(a) for some example images that are not recognized by the EFM-nearest neighbor classifier using the grayscale-LBP descriptor but are correctly recognized using the oRGB-LBP descriptor. Figure 10.8(b) shows images unrecognized using the oRGB-LBP descriptor but recognized using the CLF descriptor, Figure 10.8(c) shows images unrecognized using the CLF but recognized using the CGLF descriptor and Figure 10.8(d) shows images unrecognized using the CGLF but recognized using the CGLF+PHOG descriptor. Table 10.2 shows that for the 800 training images (100 images per class)

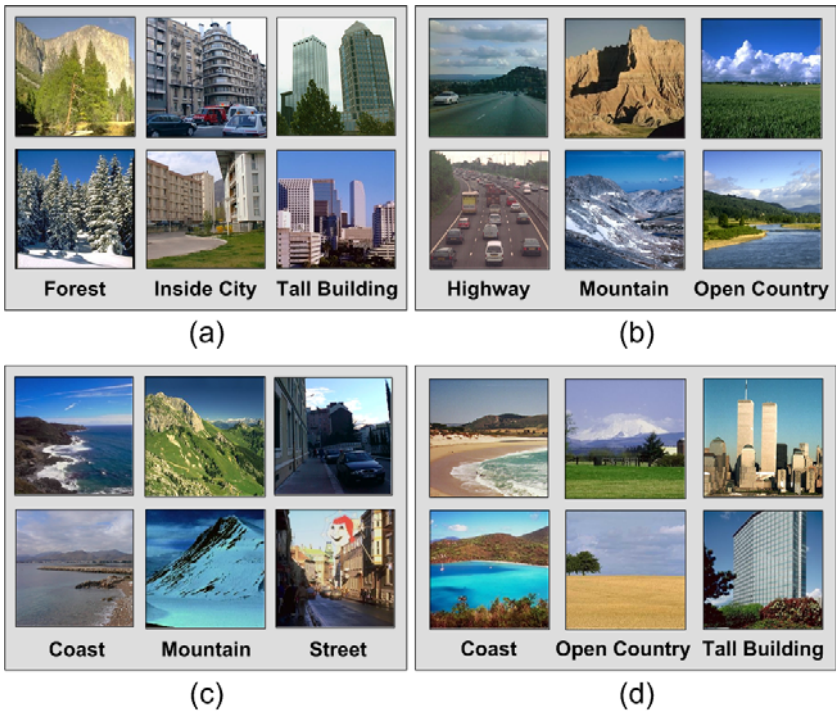

Fig. 10.8 Image recognition using the EFM-NN classifier on the MIT scene dataset: (a) example images unrecognized using the grayscale-LBP descriptor but recognized using the oRGB-LBP descriptor; (b) example images unrecognized using the oRGB-LBP descriptor but recognized using the CLF descriptor; (c) images unrecognized using the CLF but recognized using the CGLF descriptor; (d) images unrecognized using the CGLF but recognized using the CGLF+PHOG descriptor.

Table 10.2 Comparison of the Classification Performance (%) with Other Method on the MIT Scene Dataset

#train images	#test images	Our Method		Method [38]
2000	688	CLF	**86.4**	
		CGLF	**86.6**	
		CGLF+PHOG	**89.5**	
800	1888	CLF	79.3	
		CGLF	80.0	
		CGLF+PHOG	**84.3**	83.7

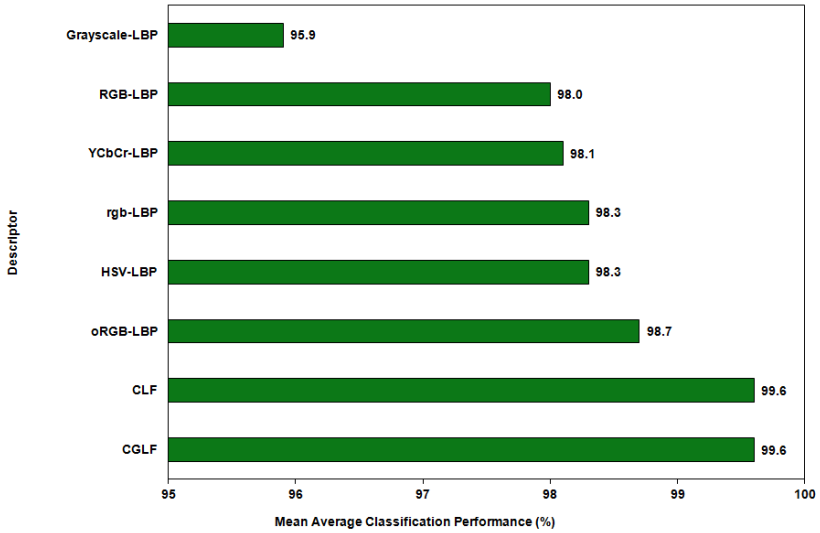

Fig. 10.9 The mean average classification performance of the eight descriptors using the EFM-NN classifier on the KTH-TIPS2-b dataset: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the CLF, and the CGLF descriptors.

and 1688 testing images the CGLF+PHOG descriptor achieves 84.3% success rate, which improves upon the result reported in [38] by 0.6%.

10.5.3 Evaluation of the Color LBP Descriptors and EFM-Nearest Neighbor Classifier on the KTH-TIPS2-b and the KTH-TIPS Datasets

We now assess the new color LBP descriptors on the KTH-TIPS2-b dataset and compare our results with those from other research groups on the same dataset. The first set of experiments assesses the overall classification performance of the eight descriptors on the KTH-TIPS2-b dataset. Note that for each category five-fold cross validation is implemented for each descriptor using the EFM-NN classifier to derive the average classification performance. Figure 10.9 shows the mean average classification performance of various descriptors. The best recognition rate that is obtained is 99.6% from the CLF and CGLF descriptors. The oRGB-LBP achieves the classification rate of 98.7%. It outperforms the other color LBP descriptors. It is noted that fusion of the color LBP descriptors (CLF) improves upon the grayscale-LBP by a significant 3.7% margin. The grayscale-LBP descriptor does not have any effect on the fusion (CGLF) result for this dataset.

Table 10.3 Category Wise Descriptor Performance (%) Split-out with the EFM-NN Classifier on the KTH-TIPS2-b Dataset. Note That the Categories are Sorted on the CGLF Results

Category	**CGLF**	**CLF**	**oRGB LBP**	**HSV LBP**	**rgb LBP**	**Gray LBP**
Aluminum Foil	**100**	**100**	**100**	**100**	**100**	**100**
Brown Bread	**100**	**100**	**100**	99	99	94
Corduroy	**100**	**100**	**100**	**100**	**100**	93
Cork	**100**	**100**	**100**	98	98	98
Cracker	**100**	**100**	96	93	93	90
Lettuce Leaf	**100**	**100**	**100**	**100**	**100**	97
Linen	**100**	**100**	**100**	99	99	99
Wood	**100**	**100**	**100**	**100**	**100**	**100**
Wool	**100**	**100**	99	**100**	**100**	96
White Bread	**99**	**99**	**99**	**99**	**99**	97
Cotton	**98**	97	97	96	96	91
Mean	**99.6**	**99.6**	**98.7**	**98.3**	**98.3**	**95.9**

The second set of experiments assesses the five best descriptors and the grayscale-LBP using the EFM-NN classifier on individual image categories. From Table 10.3 it can be seen that nine out of eleven categories achieve 100% success rate and all of the categories achieve a success rate of 98% or more with the CGLF descriptor. Aluminum Foil, Corduroy, Lettuce Leaf and Wood achieve 100% success rate across the best five descriptors. Individual color LBP features improve upon the grayscale-LBP on most of the categories. The CLF almost always improves upon the grayscale-LBP, which indicates that various color descriptors are not redundant. The CGLF very slightly improves upon the CLF. This, however, does not necessarily indicate that the grayscale information is redundant as almost all the categories show a success rate of 100% with these two descriptors. It only indicates that CLF alone contains enough information to correctly classify the texture images for the KTH-TIPS2-b dataset.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Figure 10.10(a) for some example images that are not recognized by the EFM-NN classifier using the grayscale-LBP descriptor but are correctly recognized using the oRGB-LBP descriptor. This reaffirms the importance of color and the distinctiveness of the oRGB-LBP descriptor for image category recognition. Figure 10.10(b) shows images unrecognized using the RGB-LBP descriptor but recognized using the oRGB-LBP descriptor, Figure 10.10(c) shows images unrecognized using the oRGB-LBP but recognized using the CLF descriptor, and Figure 10.10(d) shows images unrecognized using the grayscale-LBP but recognized when combined with the CLF, i.e., the CGLF descriptor.

The same set of experiments was run on the KTH-TIPS dataset with the aforementioned training and test image sets. The best result on this dataset while using a single color space was once again from the oRGB-LBP descriptor, which achieves

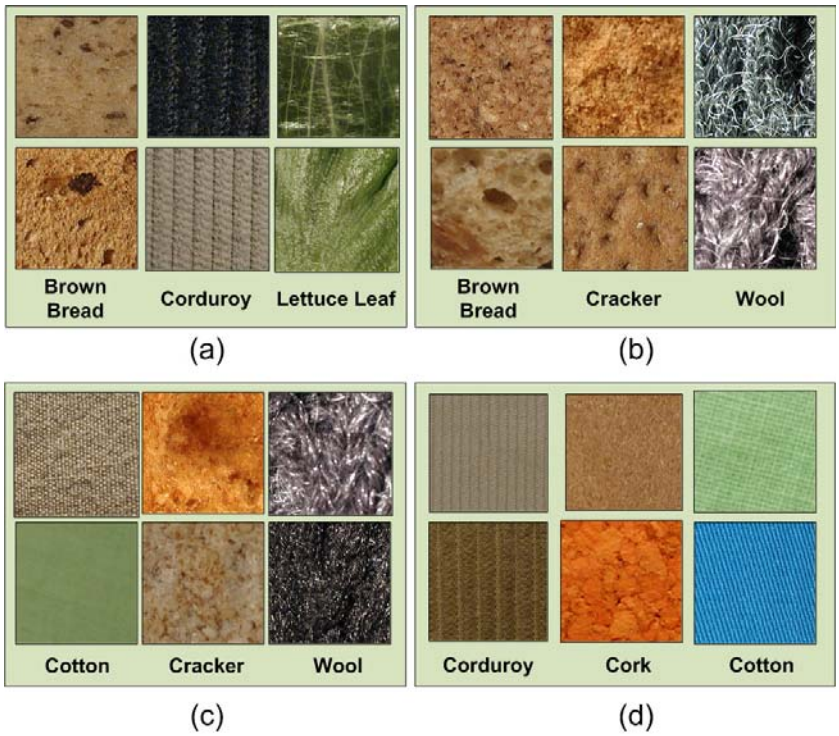

Fig. 10.10 Image recognition using the EFM-NN classifier on the KTH-TIPS2-b dataset: (a) example images unrecognized using the grayscale-LBP descriptor but recognized using the oRGB-LBP descriptor; (b) example images unrecognized using the RGB-LBP descriptor but recognized using the oRGB-LBP descriptor; (c) images unrecognized using the oRGB-LBP but recognized using the CLF descriptor; (d) images unrecognized using the grayscale-LBP but recognized using the CGLF descriptor.

Table 10.4 Comparison of the Classification Performance (%) with Other Methods on the KTH-TIPS Dataset

Methods	Performance
Our Method:	
CGLF	**99.6**
CLF	**99.6**
oRGB+LBP	**99.1**
Crosier and Griffin[10]	98.5
Kondra and Torre[19]	97.7
Zhang et. al.[49]	95.5

a 99.1% classification rate with an improvement of 3% over the grayscale-LBP. The CLF and the CGLF descriptors are tied at 99.6%. Table 10.4 shows a comparison of our results with those obtained from other methods in [10, 49, 19]. In the oRGB color space, this technique outperforms the state of the art on this dataset even without combining color descriptors. Combined LBP descriptors (CLF and CGLF) improve upon the result in [10], which was the best result on this dataset.

10.6 Conclusion

Four new color descriptors have been proposed in this chapter: the oRGB-LBP descriptor, the Color LBP Fusion (CLF), the Color Grayscale LBP Fusion (CGLF), and the CGLF+PHOG descriptors for scene image and image texture classification with applications to image search and retrieval. Experimental results using three datasets show that the oRGB-LBP descriptor improves image classification performance upon other color LBP descriptors; and the CLF, the CGLF, and the CGLF+PHOG descriptors perform better than other color LBP descriptors. The fusion of multiple Color LBP descriptors (CLF) and Color Grayscale LBP descriptor (CGLF) show improvement in image classification performance, which indicates that the various color LBP descriptors are not redundant for image classification tasks.

References

1. Agarwal, S., Roth, D.: Learning a Sparse Representation for Object Detection. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 113–127. Springer, Heidelberg (2002)
2. Banerji, S., Verma, A., Liu, C.: Novel color LBP descriptors for scene and image texture classification. In: 15th Intl. Conf. on Image Processing, Computer Vision, and Pattern Recognition, July 18-21, Las Vegas, Nevada (2011)
3. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. Computer Vision and Image Understanding 110(3), 346–359 (2008)
4. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Int. Conf. on Image and Video Retrieval, Amsterdam, The Netherlands, July 9-11, pp. 401–408 (2007)
5. Bosch, A., Zisserman, A., Munoz, X.: Scene classification using a hybrid generative/discriminative approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(4), 712–727 (2008)
6. Bratkova, M., Boulos, S., Shirley, P.: oRGB: A practical opponent color space for computer graphics. IEEE Computer Graphics and Applications 29(1), 42–55 (2009)
7. Burghouts, G., Geusebroek, J.-M.: Performance evaluation of local color invariants. Computer Vision and Image Understanding 113, 48–62 (2009)

8. Caputo, B., Hayman, E., Mallikarjuna, P.: Class-specific material categorisation. In: The Tenth IEEE International Conference on Computer Vision, Beijing, China, October 17-20, pp. 1597–1604 (2005)
9. Chen, J., Shan, S., He, C., Zhao, G., Pietikinen, M., Chen, X., Gao, W.: WLD: A robust local image descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1705–1720 (2010)
10. Crosier, M., Griffin, L.D.: Texture classification with a dictionary of basic image features. In: Proc. Computer Vision and Pattern Recognition, Anchorage, Alaska, June 23-28, pp. 1–7 (2008)
11. Datta, R., Joshi, D., Li, J., Wang, J.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40(2), 509–522 (2008)
12. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: IEEE Conf. on Computer Vision and Pattern Recognition, Madison, Wisconsin, June 16-22, vol. 2, pp. 264–271 (2003)
13. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press (1990)
14. Gevers, T., van de Weijer, J., Stokman, H.: Color feature detection: An overview. In: Lukac, R., Plataniotis, K.N. (eds.) Color Image Processing: Methods and Applications. CRC Press, University of Toronto, Ontario, Canada (2006)
15. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall (2001)
16. Grauman, K., Darrell, T.: Pyramid match kernels: Discriminative classification with sets of image features. In: Int. Conference on Computer Vision, Beijing, October 17-20, vol. 2, pp. 1458–1465 (2005)
17. Hayman, E., Caputo, B., Fritz, M., Eklundh, J.-O.: On the Significance of Real-World Conditions for Material Classification. In: Pajdla, T., Matas, J.G. (eds.) ECCV 2004, Part IV. LNCS, vol. 3024, pp. 253–266. Springer, Heidelberg (2004)
18. Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: Int. Conference on Computer Vision, Beijing, October 17-20, pp. 604–610 (2005)
19. Kondra, S., Torre, V.: Texture classification using three circular filters. In: IEEE Indian Conf. on Computer Vision, Graphics and Image Processing, Bhubaneshwar, India, December 16-19, pp. 429–434 (2008)
20. Lazebnik, S., Schmid, C., Ponce, J.: Semi-local affine parts for object recognition. In: British Machine Vision Conference, London, September 7-9, vol. 2, pp. 959–968 (2004)
21. Leung, T.K., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. Int. Journal of Computer Vision 43(1), 29–44 (2001)
22. Liu, C.: A Bayesian discriminating features method for face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(6), 725–740 (2003)
23. Liu, C.: Enhanced independent component analysis and its application to content based face image retrieval. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(2), 1117–1127 (2004)
24. Liu, C.: Gabor-based kernel PCA with fractional power polynomial models for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 572–581 (2004)
25. Liu, C.: Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5), 725–737 (2006)
26. Liu, C.: Learning the uncorrelated, independent, and discriminating color spaces for face recognition. IEEE Transactions on Information Forensics and Security 3(2), 213–222 (2008)

27. Liu, C., Wechsler, H.: Evolutionary pursuit and its application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(6), 570–582 (2000)
28. Liu, C., Wechsler, H.: Robust coding schemes for indexing and retrieval from large face databases. IEEE Transactions on Image Processing 9(1), 132–137 (2000)
29. Liu, C., Wechsler, H.: A shape and texture based enhanced Fisher classifier for face recognition. IEEE Transactions on Image Processing 10(4), 598–608 (2001)
30. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
31. Liu, C., Wechsler, H.: Independent component analysis of Gabor features for face recognition. IEEE Transactions on Neural Networks 14(4), 919–928 (2003)
32. Liu, C., Yang, J.: ICA color space for pattern recognition. IEEE Transactions on Neural Networks 20(2), 248–257 (2009)
33. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60(2), 91–110 (2004)
34. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
35. Niblack, W., Barber, R., Equitz, W.: The QBIC project: Querying images by content using color, texture and shape. In: SPIE Conference on Geometric Methods in Computer Vision II, pp. 173–187 (1993)
36. Ojala, T., Pietikainen, M., Harwood, D.: Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: Int. Conf. on Pattern Recognition, Jerusalem, Israel, pp. 582–585 (1994)
37. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1), 51–59 (1996)
38. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. Journal of Computer Vision 42(3), 145–175 (2001)
39. Pontil, M., Verri, A.: Support vector machines for 3D object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(6), 637–646 (1998)
40. Schiele, B., Crowley, J.: Recognition without correspondence using multidimensional receptive field histograms. Int. Journal of Computer Vision 36(1), 31–50 (2000)
41. Shih, P., Liu, C.: Comparative assessment of content-based face image retrieval in different color spaces. International Journal of Pattern Recognition and Artificial Intelligence 19(7), 873–893 (2005)
42. Smith, A.R.: Color gamut transform pairs. Computer Graphics 12(3), 12–19 (1978)
43. Stokman, H., Gevers, T.: Selection and fusion of color models for image feature detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), 371–381 (2007)
44. Swain, M.J., Ballard, D.H.: Color indexing. International Journal of Computer Vision 7(1), 11–32 (1991)
45. Verma, A., Banerji, S., Liu, C.: A new color SIFT descriptor and methods for image category classification. In: International Congress on Computer Applications and Computational Science, Singapore, December 4-6, pp. 819–822 (2010)
46. Verma, A., Liu, C.: Novel EFM-KNN classifier and a new color descriptor for image classification. In: 20th IEEE Wireless and Optical Communications Conference (Multimedia Services and Applications), Newark, New Jersey, USA, April 15-16 (2011)

47. Weber, M., Welling, M., Perona, P.: Towards automatic discovery of object categories. In: IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC, June 13-15, vol. 2, pp. 2101–2109 (2000)
48. Yang, J., Liu, C.: Color image discriminant models and algorithms for face recognition. IEEE Transactions on Neural Networks 19(12), 2088–2098 (2008)
49. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. Int. Journal of Computer Vision 73(2), 213–238 (2007)
50. Zhu, C., Bichot, C., Chen, L.: Multi-scale color local binary patterns for visual object classes recognition. In: Int. Conf. on Pattern Recognition, Istanbul, Turkey, August 23-26, pp. 3065–3068 (2010)