Novel Color LBP Descriptors for Scene and Image Texture Classification

Sugata Banerji[1], Abhishek Verma and Chengjun Liu

***Abstract*—Four novel color Local Binary Pattern (LBP) descriptors are presented in this paper 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 grand challenge 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.**

***Index Terms*—The oRGB-LBP descriptor, the Color LBP Fusion (CLF) descriptor, the Color Grayscale LBP Fusion (CGLF) descriptor, the CGLF+PHOG descriptor, Enhanced Fisher Model (EFM), image search.**

I. 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 significantly larger amount of discriminative information [1]-[4]. Color based image search can be very useful in the identification of object and natural scene categories [4]. 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 [5], [6].

In recent years, the recognition and classification of textures using the Local Binary Pattern (LBP) features has been shown to be promising [7]-[11]. Color features when combined with the intensity based texture descriptors are able to outperform many alternatives. In this paper, we employ a variable mask size to generate a multi-scale LBP feature vector that is more robust to changes of scale and orientation. Further, we extend the multi-scale LBP descriptor to different color spaces including the recently proposed oRGB color space [12] and propose a new multi-scale oRGB-LBP feature representation, and then integrate it 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 PHOG to obtain the novel CGLF+PHOG descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) [13], [14] and image classification is based on the nearest neighbor classification rule (EFM-NN). The effectiveness of the proposed descriptors and classification method will be evaluated using three grand challenge datasets: the MIT scene database, the KTH-TIPS2-b and the KTH-TIPS materials databases.

The rest of the paper is structured as follows: In Section II we briefly overview several representative methods on color image representation, search and retrieval. Section III describes the new color descriptors and presents an overview of five color spaces. In Section IV we present the details of the EFM feature extraction technique. Section V describes the datasets used in our experiments along with a detailed evaluation of color descriptors and classification methodology. Finally, conclusions are drawn in Section VI.

II. Related Work

In this section we briefly survey representative work on color image search and retrieval. In recent years, use of color as a means to face recognition [1], [3], [15] and object and scene retrieval [4] has gained popularity. Color features can capture discriminative information by means of the color invariants, color histogram, color texture, etc. One of the earlier works in this field is the color indexing system designed by Swain and Ballard, which uses the color

[1] Corresponding author

S. Banerji, A. Verma and C. Liu are with the Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102 USA (e-mail: {sb256, av56, chengjun.liu}@njit.edu).

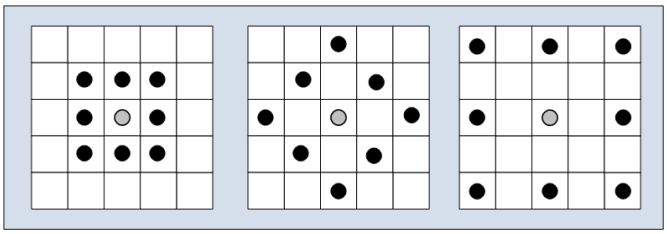

Fig. 1. The multi-scale LBP operators.

histogram for image inquiry from a large image database [16]. More recent work on color based image search appears in [1], [2], [4], [17] that propose several new color spaces and methods for face, object and scene category recognition. Evaluation of local color invariant descriptors is performed in [5]. Fusion of color models, color region detection and color edge detection have been investigated for representation of color images [6]. Key contributions in color, texture, and shape abstraction have been discussed in Datta et al. [18].

Lately, several methods based on LBP features have been proposed for image representation and classification [9], [11]. Extraction of LBP features is computationally efficient and with the use of multi-scale filters; invariance to scaling and rotation can be achieved [9]. Fusion of different features has been shown to achieve a good retrieval success rate [11], [19]. Local image descriptors have also been shown to perform well for texture based image retrieval [10], [19].

Efficient retrieval requires a robust feature extraction method that has the ability to learn meaningful low-dimensional patterns in spaces of very high dimensionality [20]-[22]. Low-dimensional representations are also important when one considers the intrinsic computational aspect. PCA has been widely used to perform dimensionality reduction for image indexing and retrieval [13], [23]. Recently, Support Vector Machine (SVM) classifier for multiple category recognition has gained popularity [19] though it suffers from the drawback of being computationally too expensive on large scale image classification tasks. The EFM feature extraction method has achieved good success for the task of image based representation and retrieval [14], [24], [25].

III. New Color LBP Descriptors

We first review in this section five color spaces in which our four novel color descriptors are defined, and then discuss the conventional LBP descriptor which is the intensity-based (grayscale) LBP descriptor. Next we present the new oRGB-LBP descriptor and two novel combined color multi-scale LBP descriptors: the Color LBP Fusion (CLF), and the Color Grayscale LBP Fusion (CGLF). Finally, we present the new CGLF+PHOG descriptor obtained from combining the CGLF with the Pyramid of Histograms of Orientation Gradients.

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

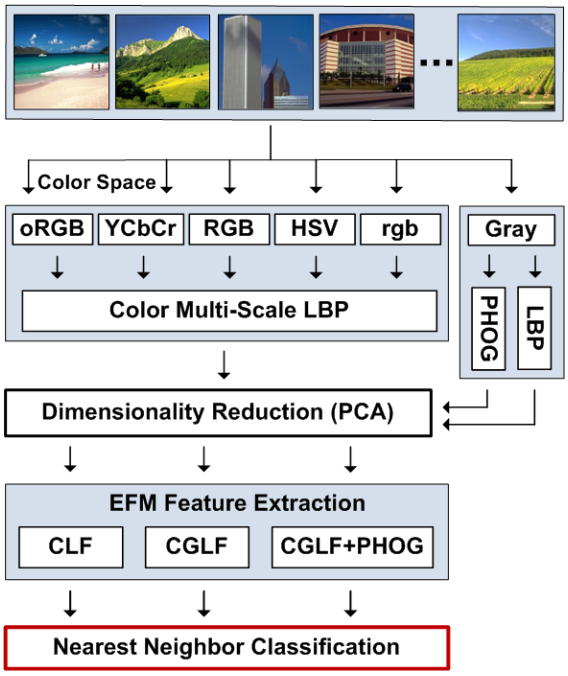

Fig. 2. An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

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. The HSV color space is motivated by 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 [26]. The YCbCr color space is developed for digital video

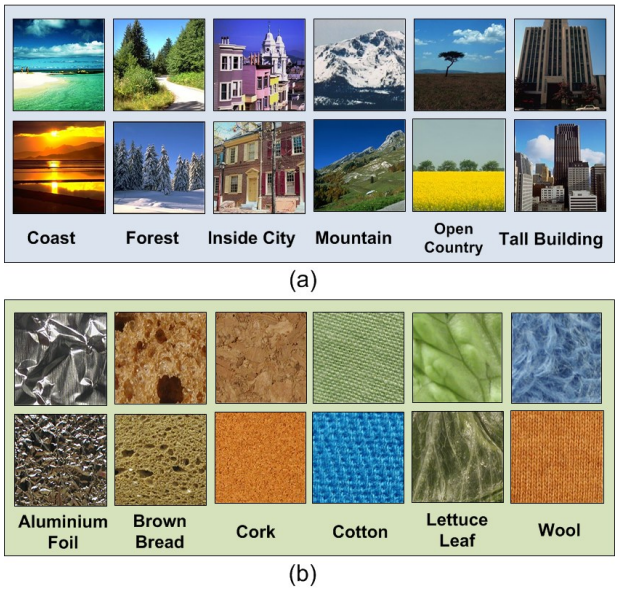

Fig. 3. Example images from (a) the MIT scene dataset and (b) the KTH-TIPS2-b materials dataset.

TABLE I
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**

standard and television transmissions. In YCbCr, the RGB components are separated into luminance, chrominance blue, and chrominance red.

The oRGB color space [12] has three channels L, C_1 and C_2. 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 C_1 and C_2. The value of C_1 lies within [-1, 1] and the value of C_2 lies within [-0.8660, 0.8660]. The L channel contains the luminance information and its values range between [0, 1].

$$\begin{bmatrix} L \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

The LBP descriptor proposed by Ojala et al. [7], [8] assigns an intensity value to each pixel of an image based on the intensity values of the 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 [9]. Using the multi-scale LBP operator shown in Fig. 1, we generate three LBP images 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 from concatenating the 768 dimensional vectors from the three channels. As a result, the four color LBP descriptors are defined: the RGB-LBP, the YCbCr-LBP, the HSV-LBP, and the rgb-LBP descriptors.

The Pyramid of Histograms of Orientation Gradients (PHOG) descriptor [27] 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 [27].

The four new color LBP descriptors are defined in the oRGB color space and the fusion in different color spaces. 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 further the CLF descriptor and the grayscale-LBP descriptor. The CGLF is combined with the PHOG to obtain the CGLF+PHOG descriptor.

IV. THE EFM FEATURE EXTRACTION METHOD AND NEAREST NEIGHBOR CLASSIFICATION

Image classification using the new descriptor introduced in the preceding section is implemented using the Enhanced Fisher Model (EFM) feature extraction method [13], [14] and the nearest neighbor classification rule (EFM-NN).

Let $X \in \mathrm{R}^N$ be a random vector whose covariance matrix is Σ_X:

$$\Sigma_X = \varepsilon\{[X - \varepsilon(X)][X - \varepsilon(X)]^t\} \quad (2)$$

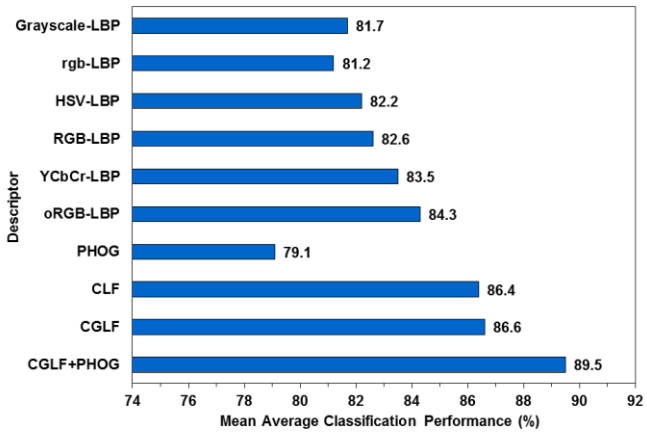

Fig. 4. The mean average classification performance of the ten descriptors using the EFM-NN classifier on the MIT scene dataset: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the PHOG, the CLF, the CGLF, and the CGLF+PHOG descriptors.

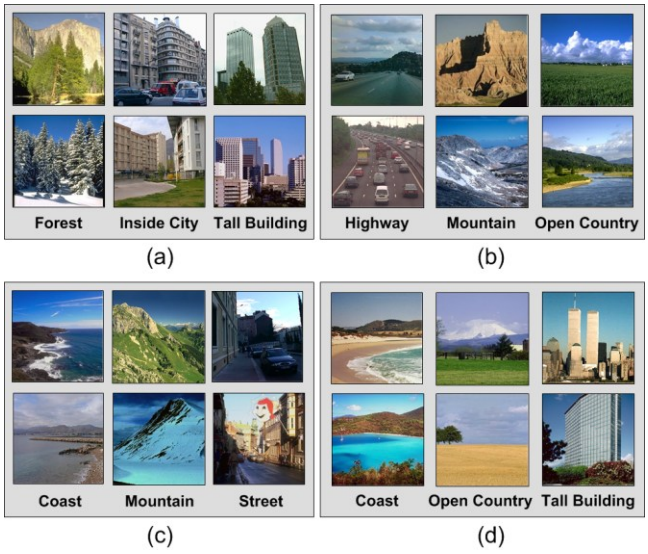

Fig. 5. Image recognition using the EFM-NN classifier on the MIT scene dataset: (a) example images unrecognized using the grayscale-LBP descriptor but recognized using the oRGB-LBP descriptor; (b) example images unrecognized using the oRGB-LBP descriptor but recognized using the CLF descriptor; (c) images unrecognized using the CLF but recognized using the CGLF descriptor; (d) images unrecognized using the CGLF but recognized using the CGLF+PHOG descriptor.

where $\varepsilon(.)$ is the expectation operator and t denotes the transpose operation. The eigenvectors of the covariance matrix Σ_X can be derived by PCA:

$$\Sigma_X = \Phi \Lambda \Phi^t \tag{3}$$

where $\Phi = [\phi_1 \phi_2 \phi_3 .. \phi_N]$ is an orthogonal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_N\}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order. An important application of PCA is dimensionality reduction:

$$Y = P^t X \tag{4}$$

where $P = [\phi_1 \phi_2 \phi_3 .. \phi_K]$, and $K < N$. $X \in \mathrm{R}^K$ 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 [28] 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 [28]:

$$S_w = \sum_{i=1}^{L} P(\omega_i) \varepsilon\{(Y - M_i)(Y - M_i)^t \mid \omega_i\} \tag{5}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M_0)(M_i - M_0)^t \tag{6}$$

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 $S_w^{-1} S_b$ [28]:

$$S_w^{-1} S_b \Psi = \Psi \Delta \tag{7}$$

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 Y onto the eigenvectors of Ψ:

$$Z = \Psi^t Y \tag{8}$$

Z thus is more effective than the feature vector 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 [13]. A new method, the Enhanced Fisher Model (EFM), is capable of improving the generalization performance of the FLD method [13]. 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 [13]. The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by Fukunaga [28]: whitening the within-class scatter matrix and applying PCA to 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 [13], 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 (in the reduced PCA space) are not too small (for better generalization performance) [13].

TABLE II
COMPARISON OF THE CLASSIFICATION PERFORMANCE (%) WITH OTHER METHOD ON THE MIT SCENE DATASET

# train	# test	Our Method		[29]
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

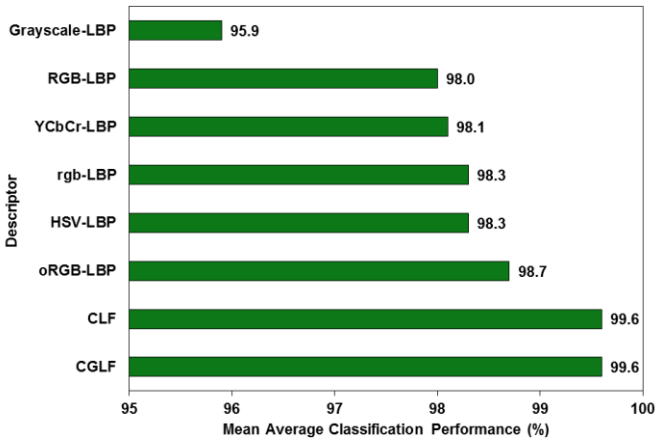

Fig. 6. The mean average classification performance of the eight descriptors using the EFM-NN classifier on the KTH-TIPS2-b dataset: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the CLF, and the CGLF descriptors.

Image classification is implemented using the nearest neighbor classification rule. Fig. 2 gives an overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

V. Experiments

A. Datasets and Experimental Methodology

We apply the following three publicly accessible datasets to evaluate our proposed descriptors and classification method: the MIT Scene dataset [29], the KTH-TIPS (Textures under varying Illumination, Pose and Scale) [11] and KTH-TIPS2-b datasets [30]. The MIT scene dataset [29] has 2,688 images classified as eight categories: 360 coast, 328 forest, 374 mountain, 410 open 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) [29]. See Fig. 3(a) for some sample images from this dataset. The KTH-TIPS dataset [11], [30], [31] 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 which makes this dataset moderately challenging. The KTH-TIPS2-b dataset [30] 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 Fig. 3(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 KTH-TIPS2-b dataset we use five random sets of 200 training images per class and 100 testing images per class. For the KTH-TIPS dataset we select five random sets of 40 training images per class and 41 test images per class (same numbers as used in [11], [19], [31]). For the MIT scene dataset we randomly select five sets and 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.

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

The first set of experiments on this dataset assesses the overall classification performance of the ten descriptors. Note that for each category we implement five-fold cross validation for each descriptor using the EFM-nearest neighbor classifier to derive the average classification performance. As a result, each descriptor yields 8 average classification 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. Fig. 4 shows the mean average classification performance of various descriptors. The best recognition rate that we obtain is 89.5% from the CGLF+PHOG, which is a very respectable value 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

TABLE III
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
Aluminium 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**

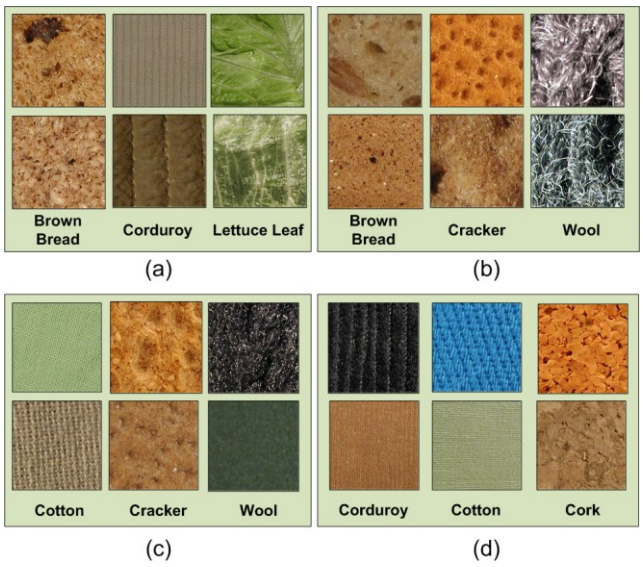

Fig. 7. Image recognition using the EFM-NN classifier on the KTH-TIPS2-b dataset: (a) example images unrecognized using the grayscale-LBP descriptor but recognized using the oRGB-LBP descriptor; (b) example images unrecognized using the RGB-LBP descriptor but recognized using the oRGB-LBP descriptor; (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.

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 I 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 Fig. 5(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. Fig. 5(b) shows images unrecognized using the oRGB-LBP descriptor but recognized using the CLF descriptor, Fig. 5(c) shows images unrecognized using the CLF but recognized using the CGLF descriptor and Fig. 5(d) shows images unrecognized using the CGLF but recognized using the CGLF+PHOG descriptor.

From Table II it can be seen that on the 800 training images (100 images per class) and 1688 testing images we achieve 84.3% success rate with the CGLF+PHOG descriptor. This improves over the result of authors in [29] by 0.6%.

C. Evaluation of Novel Color Descriptors and EFM-Nearest Neighbor Classifier on the KTH-TIPS2-b and the KTH-TIPS Datasets

Here we present a detailed experimental evaluation on the KTH-TIPS2-b dataset followed by a comparison of success rate with other research groups on the KTH-TIPS 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 we implement five-fold cross validation for each descriptor using the EFM-nearest neighbor classifier to derive the average classification performance. Fig. 6 shows the mean average classification performance of various descriptors. The best recognition rate that we obtain 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 in case of this dataset.

The second set of experiments assesses the five best descriptors and the grayscale-LBP using the EFM-nearest neighbor classifier on individual image categories. From Table III 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. Aluminium 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, this 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 in the case of KTH-TIPS2-b dataset.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Fig. 7(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. This reaffirms the importance of color and the distinctiveness of the oRGB-LBP

TABLE IV
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 [11]	98.5
Kondra and Torre [31]	97.7
Zhang et al. [19]	95.5

descriptor for image category recognition. Fig. 7(b) shows images unrecognized using the RGB-LBP descriptor but recognized using the oRGB-LBP descriptor, Fig. 7(c) shows images unrecognized using the oRGB-LBP but recognized using the CLF descriptor, and Fig. 7(d) shows images unrecognized using the grayscale-LBP but recognized when combined with the CLF, i.e., the CGLF descriptor.

We ran the same set of experiments 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 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 IV shows a comparison of our results with those obtained from other methods in [11], [19], [31]. In the oRGB color space, our 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 [11], previously the best result on this dataset.

VI. Conclusion

We have proposed four new color descriptors: the oRGB-LBP descriptor, and then integrated it with other color LBP features to produce 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. Results of the experiments using three grand challenge datasets show that our oRGB-LBP descriptor improves recognition performance upon other color LBP descriptors, the CLF, the CGLF, and the CGLF+PHOG descriptors perform better than the other color LBP descriptors. The fusion of multiple Color LBP descriptors (CLF) and Color Grayscale LBP descriptor (CGLF) show improvement in the classification performance, which indicates that various color LBP descriptors are not redundant for image classification tasks.

References

[1] C. Liu and J. Yang, "ICA color space for pattern recognition," *IEEE Trans. on Neural Networks*, vol. 20, no. 2, pp. 248-257, 2009.

[2] J. Yang and C. Liu, "Color image discriminant models and algorithms for face recognition," *IEEE Transactions on Neural Networks*, vol. 19, no. 12, pp. 2088-2098, 2008.

[3] P. Shih and C. Liu, "Comparative assessment of content-based face image retrieval in different color spaces," *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 873-893, 2005.

[4] A. Verma, S. Banerji, and C. Liu, "A new color SIFT descriptor and methods for image category classification," *in the 2010 Int. Congress on Computer Applications and Computational Science*, pp. 819-822, Singapore.

[5] G. Burghouts and J.M. Geusebroek, "Performance evaluation of local color invariants," *Computer Vision and Image Understanding*, vol. 113, pp. 48-62, 2009.

[6] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 371-381, 2007.

[7] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," *in the 1994 Int. Conf. on Pattern Recognition,* pp. 582-585, Jerusalem, Israel.

[8] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[9] C. Zhu, C. Bichot, and L. Chen, "Multi-scale color local binary patterns for visual object classes recognition," *in the 2010 Int. Conf. on Pattern Recognition*, pp. 3065-3068, Istanbul, Turkey.

[10] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705-1720, 2010.

[11] M. Crosier and L.D. Griffin, "Texture classification with a dictionary of basic image features," *in the 2008 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-7, Anchorage, Alaska.

[12] M. Bratkova, S. Boulos, and P. Shirley, "oRGB: a practical opponent color space for computer graphics," *IEEE Computer Graphics and Applications*, vol. 29, no. 1, pp. 42-55, 2009.

[13] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132-137, 2000.

[14] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[15] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725-737, 2006.

[16] M.J. Swain and D.H. Ballard, "Color indexing," *Int. Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[17] C. Liu, "Learning the uncorrelated, independent, and discriminating color spaces for face recognition," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 2, pp. 213-222, 2008.

[18] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 509-522, 2008.

[19] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories a comprehensive study," *Int. Journal of Computer Vision*, vol. 73, no. 2, pp. 213-238, 2007.

[20] C. Liu, "A bayesian discriminating features method for face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 725-740, 2003.

[21] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000.

[22] C. Liu and H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 919-928, 2003.

[23] C. Liu, "Gabor-based kernel with fractional power polynomial models for face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572-581, 2004.

[24] C. Liu and H. Wechsler, "A shape and texture based enhanced fisher classifier for face recognition," *IEEE Trans. on Image Processing*, vol. 10, no. 4, pp. 598-608, 2001.

[25] C. Liu, "Enhanced independent component analysis and its application to content based face image retrieval," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 2, pp. 1117-1127, 2004.

[26] C.G. Gonzalez and R.E. Woods, *Digital Image Processing*, Upper Saddle River, NJ: Prentice-Hall, 2001.

[27] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," *in the 2007 Int. Conf. on Image and Video Retrieval*, pp. 401-408, New York, NY.

[28] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, San Diego, CA: Academic Press, 2nd ed., 1990.

[29] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *Int. Journal of Computer Vision*, vol. 42, no. 3, pp. 145-175, 2001.

[30] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," *in the 2005 IEEE Int. Conf. on Computer Vision,* pp. 1597-1604, Beijing, China.

[31] S. Kondra and V. Torre, "Texture classification using three circular filters," *in the 2008 Sixth IEEE Indian Conf. on Computer Vision, Graphics & Image Processing,* pp. 429-434, Bhubaneshwar, India.