Novel Color HWML Descriptors for Scene and Object Image Classification

Sugata Banerji[1], Atreyee Sinha and Chengjun Liu
Department of Computer Science,
New Jersey Institute of Technology,
Newark, NJ 07102, USA
Email:{sb256, as739, chengjun.liu}@njit.edu

***Abstract*—Several new image descriptors are presented in this paper that combine color, texture and shape information to create feature vectors for scene and object image classification. In particular, first, a new three dimensional Local Binary Patterns (3D-LBP) descriptor is proposed for color image local feature extraction. Second, three novel color HWML (HOG of Wavelet of Multiplanar LBP) descriptors are derived by computing the histogram of the orientation gradients of the Haar wavelet transformation of the original image and the 3D-LBP images. Third, the Enhanced Fisher Model (EFM) is applied for discriminatory feature extraction and the nearest neighbor classification rule is used for image classification. Finally, the Caltech 256 object categories database and the MIT scene dataset are used to show the feasibility of the proposed new methods.**

***Index Terms*—The HOG of Wavelet of Multiplanar LBP (HWML) descriptor, the Fused Color HWML (FCHWML) descriptor, Enhanced Fisher Model (EFM), image search, scene and object classification**

I. Introduction

In recent years, use of color as a means to image retrieval [1], [2] and object and scene search [3], [4] has gained popularity. 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 [5], [6], [7]. 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 [8]. 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.

More recent work on color based image classification appears in [9], [3], [10] 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 [11], and the evaluation of local color invariant descriptors is performed in [12]. Fusion of color models, color region detection and color edge detection has been investigated for representation of color images [13]. Some important contributions of color, texture, and shape abstraction for image retrieval have been discussed in Datta et al. [14].

[1]Corresponding author

Lately, several methods based on LBP features have been proposed for image representation and classification [15], [16]. 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 LBP features is computationally efficient and with the use of multi-scale filters invariance to scaling and rotation can be achieved [15], [4]. Fusion of different features has been shown to achieve a good image retrieval success rate [4], [16], [17]. Local image descriptors have also been shown to perform well for texture based image retrieval [4], [17]. Several researchers have used the Haar wavelet transform for object detection in images and LBP has also been combined with Haar-like features for face detection [18]. The Histogram of Orientation Gradients (HOG) descriptor [19], [20] is able to represent an image by its local

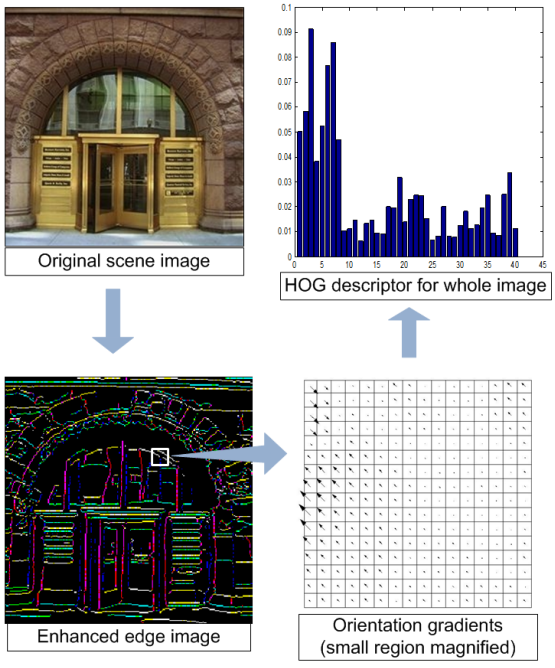

Fig. 1. The Histogram of Orientation Gradients (HOG) descriptor.

978-1-4673-2584-4/12/$31.00 ©2012 IEEE

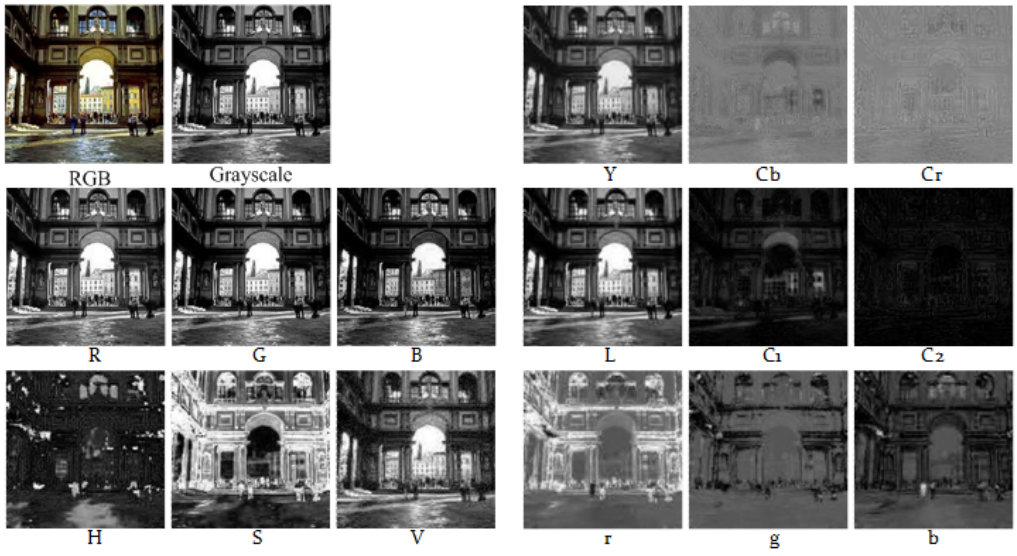

Fig. 2. A sample image from the MIT Scene dataset split up into various color components and grayscale.

shape and the spatial layout of the shape. The local shape is captured by the distribution over edge orientations within a region. Figure 1 shows how the HOG descriptor is formed by the gradient histograms from a scene image.

Efficient retrieval requires a robust feature extraction method that has the ability to learn meaningful low-dimensional patterns in spaces of very high dimensionality [21], [22], [23]. Low-dimensional representation is also important when one considers the intrinsic computational aspect. PCA has been widely used to perform dimensionality reduction for image indexing and retrieval [24], [25]. The EFM feature extraction method has achieved good success for the task of image based representation and retrieval [26].

In this paper, we employ three masks in three perpendicular planes to generate a novel multiplanar 3D-LBP feature that contains more information than the traditional LBP. Further, we subject the 3D-LBP image to a Haar wavelet transformation and then generate the HOG of the resultant image to create a robust feature vector. We extend this concept to different color spaces and propose the new oRGB-HWML and YCbCr-HWML feature representations, and then integrate them with other color HWML features to produce the novel Fused Color HWML (FCHWML) descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) [24] and image classification is based on the nearest neighbor classification rule.

II. Implementation Details

We first review in this section five color spaces in which our new descriptor is defined, and then discuss the 3D-LBP descriptor which is an improvement upon the traditional

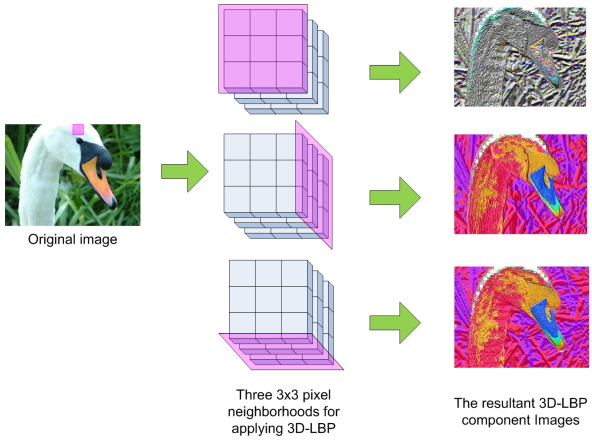

Fig. 3. The proposed 3D-LBP descriptor. A 3x3x3 pixel region of the original image is magnified to show the 3D-LBP neighborhoods and the resulting LBP images.

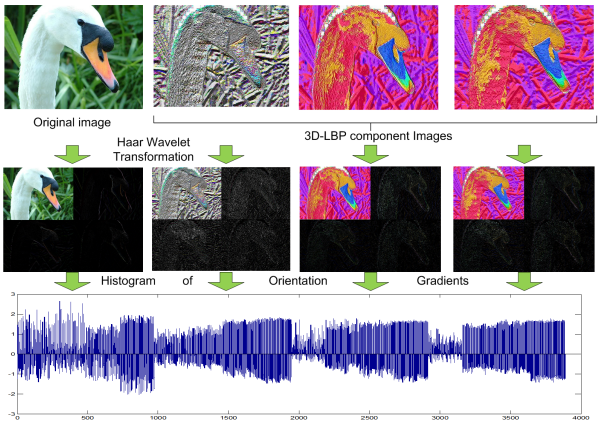

Fig. 4. The proposed HWML descriptor. The original and 3D-LBP images undergo Haar Wavelet Transformation and then HOG is generated for each component of the resulting image and concatenated.

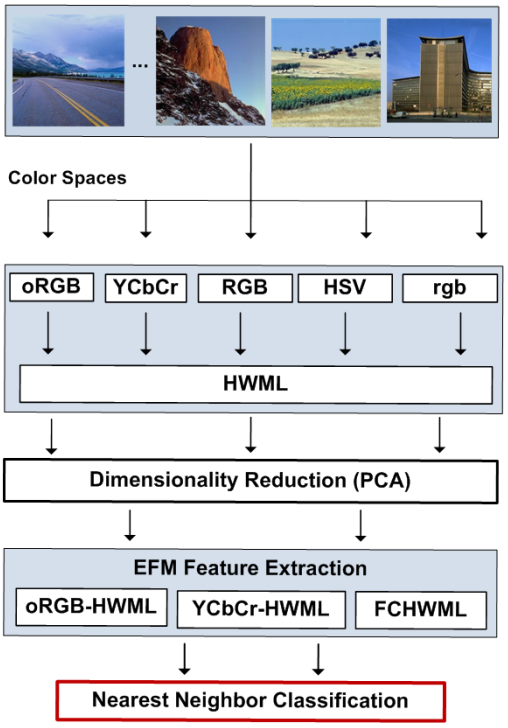

Fig. 5. An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

LBP descriptor. Next we present the new HWML descriptor in oRGB and YCbCr color spaces and a combined color FCHWML descriptor.

A. Color spaces

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. 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 [27]. 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:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

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

The oRGB color space [28] 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 value of $C1$ lies within $[-1, 1]$ and the value of $C2$ lies within $[-0.8660, 0.8660]$. The L channel contains the luminance information and its values range between $[0, 1]$:

$$\begin{bmatrix} L \\ C1 \\ C2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

The five color spaces used by us in this paper are shown in Figure 2. The grayscale image has also been included as a reference.

B. The 3D-LBP and HWML descriptors

The LBP descriptor assigns an intensity value to each pixel of an image based on the intensity values of the eight neighboring pixels. Since a color image is represented by a three dimensional matrix, we extend this concept to assign an intensity value to each pixel based on its neighboring pixels not only on the same color plane but on other planes as well. This method is explained in Figure 3. We replicate the first and third image planes on opposite sides of the three existing planes to create a five-plane matrix. After the LBP operation, only the three middle planes are retained.

The 3D-LBP method produces three images. We then apply the Haar wavelet transformation to the original and these three images to divide each image into four distinct regions. We then generate the HOG descriptor for each of these regions of the four images and then concatenate them to get our final HWML feature vector. This process is illustrated in Figure 4.

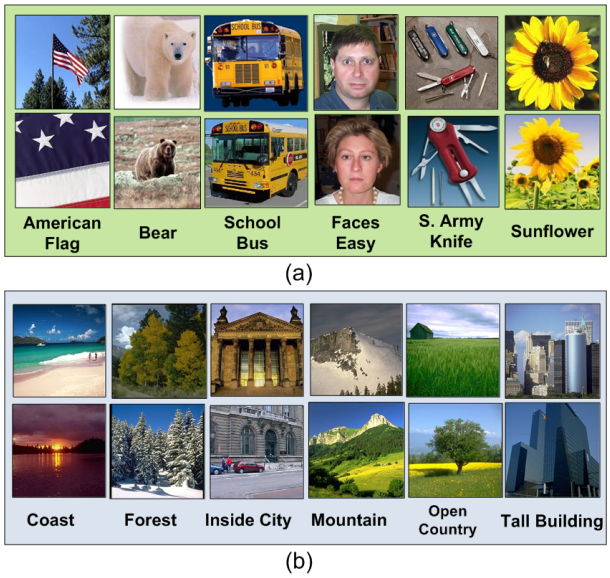

Fig. 6. Example images from the (a) Caltech 256 dataset and (b) the MIT Scene dataset.

TABLE I
CATEGORY WISE DESCRIPTOR PERFORMANCE (%) ON THE TOP 15 CLASSES OF THE CALTECH 256 DATASET. NOTE THAT THE CATEGORIES ARE SORTED ON THE FCHWML RESULTS

Category	Fusion	YCbCr	oRGB	RGB	HSV
car-side	**100**	**100**	**100**	**100**	**100**
faces-easy	**100**	**100**	**100**	**100**	**100**
airplanes	**100**	92	96	92	**100**
motorbikes	**100**	88	96	88	**100**
bonsai	**100**	88	92	80	68
leopards	**96**	**96**	**96**	88	**96**
sunflower	92	**100**	96	88	80
hibiscus	**92**	88	**92**	80	**92**
watch	**88**	84	**88**	76	**88**
grand-piano	**88**	76	80	**88**	76
ketch	84	76	76	84	**88**
telephone-box	**84**	72	76	52	56
american-flag	**80**	60	72	44	40
school-bus	76	**80**	**80**	72	**80**
tomato	**76**	56	52	48	28

TABLE II
COMPARISON OF THE CLASSIFICATION PERFORMANCE (%) WITH OTHER METHODS ON CALTECH 256 DATASET

#train	#test	Proposed Method		[3]	
15360	5120	oRGB-HWML	**30.7**		
		YCbCr-HWML	**31.1**		
		FCHWML	**34.7**		
12800	6400	oRGB-HWML	**29.0**	oRGB-SIFT	23.9
		YCbCr-HWML	**29.7**	CSF	30.1
		FCHWML	**33.0**	CGSF	**35.6**

C. The EFM-NN Classifier

Image classification using the new descriptor introduced in the preceding section is implemented using the Enhanced Fisher Model (EFM) feature extraction method [24] and the Nearest Neighbor classification rule (EFM-NN).

Let $\mathscr{X} \in \mathbb{R}^N$ be a random vector whose covariance matrix is $\Sigma_{\mathscr{X}}$:

$$\Sigma_{\mathscr{X}} = \mathscr{E}\{[\mathscr{X} - \mathscr{E}(\mathscr{X})][\mathscr{X} - \mathscr{E}(\mathscr{X})]^t\} \tag{3}$$

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 PCA:

$$\Sigma_{\mathscr{X}} = \Phi \Lambda \Phi^t \tag{4}$$

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 diagonal elements in decreasing order. An important application of PCA is dimensionality reduction:

$$\mathscr{Y} = P^t \mathscr{X} \tag{5}$$

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 [29] 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 [29]:

$$S_w = \sum_{i=1}^{L} P(\omega_i) \mathscr{E}\{(\mathscr{Y} - M_i)(\mathscr{Y} - M_i)^t | \omega_i\} \tag{6}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t \tag{7}$$

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 $S_w^{-1} S_b$ [29]:

$$S_w^{-1} S_b \Psi = \Psi \Delta \tag{8}$$

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 of Ψ:

$$\mathscr{Z} = \Psi^t \mathscr{Y} \tag{9}$$

$\mathscr{Z}$ 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 [24]. A better method, the Enhanced Fisher Model (EFM), is capable of improving the generalization performance of the FLD method [24]. 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 [24]. The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by [29]: 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 [24], 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) [24].

Image classification is implemented using the nearest neighbor classification rule. Figure 5 gives an overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

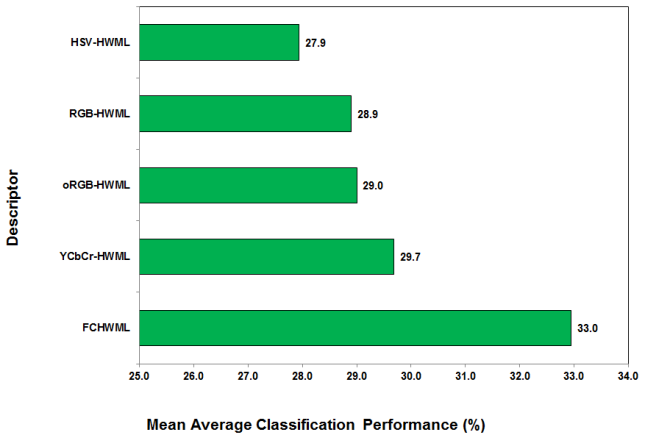

Fig. 7. The mean average classification performance of the five descriptors using the EFM-NN classifier on the Caltech 256 dataset: the RGB-HWML, the HSV-HWML, the oRGB-HWML, the YCbCr HWML and the FCHWML descriptors.

III. Experimental results

A. Caltech 256 Dataset

The Caltech 256 dataset [30] holds 30,607 images divided into 256 object categories and a clutter class. The images have high intra-class variability and high object location variability. Each category contains at least 80 images, at most 827 images and 119 images on average. The images represent a diverse set of lighting conditions, poses, backgrounds, and sizes. Images are in color, in JPEG format with only a small percentage in grayscale. The average size of each image is 351x351 pixels. Some sample images from this dataset can be seen in Figure 6 (a).

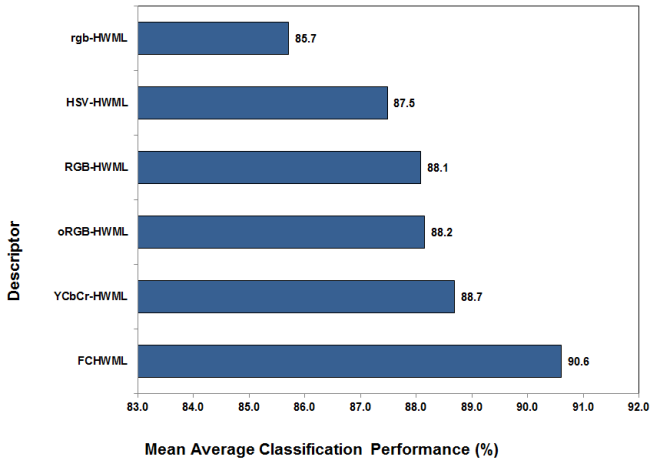

Fig. 8. The mean average classification performance of the six descriptors using the EFM-NN classifier on the MIT Scene dataset: the rgb-HWML, the HSV-HWML, RGB-HWML, the oRGB-HWML, the YCbCr HWML and the FCHWML descriptors.

TABLE III
Category wise descriptor performance (%) on the MIT Scene dataset. Note that the categories are sorted on the FCHWML results

Category	Fusion	YCbCr	oRGB	RGB	HSV	rgb
forest	**98**	**98**	96	97	97	95
coast	**94**	90	92	93	90	91
street	94	88	**95**	89	89	88
inside city	**92**	**92**	**92**	**92**	88	89
mountain	**91**	90	90	89	88	86
tall building	**90**	89	86	87	88	86
highway	**88**	**88**	**88**	86	86	84
open country	**78**	75	76	71	74	68
Mean	**90.6**	**88.7**	**88.2**	**88.1**	**87.5**	**85.7**

On this dataset, we conduct experiments for HWML descriptors from four different color spaces and their fusion. For each class, we make use of 50 images for training and 25 images for testing. The data splits are the ones that are provided on the Caltech website [30]. Figure 7 shows the detailed performance of our EFM-NN classification technique on this dataset. The best recognition rate that we obtain is 33.0%, which is a very respectable value for a dataset of this size and complexity. It can be seen from our previous work [3] that dense histograms perform poorly on this dataset as the intra-class variability is very high and in several cases the object occupies a small portion of the full image. Also, previously processor-intensive SIFT and visual vocabulary based methods achieved the best classification rate of 23.9% for a single color space and 35.6% after color and grayscale fusion. The proposed method is faster than the SIFT-based method and the YCbCr-HWML alone achieves a success rate of 29.7%. Fusion of color spaces improves our result further by over 3%. Due to the nature of the 3D-LBP descriptor, it is not defined for grayscale images and so we did not conduct any experiments for grayscale. Also, conversion to the rgb color space is undefined for grayscale images and we did not use the rgb color space on this dataset as it contains some grayscale images.

Table II compares our results with those of SIFT-based methods. Table I shows the descriptor performance for the top 15 categories from this dataset. The FCHWML recognition rates for the top 15 categories lie between 76% and 100% with five categories having full success rate.

TABLE IV
Comparison of the Classification Performance (%) with Other Methods on the MIT Scene Dataset

#train	#test	HWML		[4]		[31]
2000	688	oRGB	**88.2**	CLF	**86.4**	-
		YCbCr	**88.7**	CGLF	**86.6**	
		Fused	**90.6**	CGLF+PHOG	**89.5**	
800	1888	oRGB	**85.8**	CLF	79.3	
		YCbCr	**85.7**	CGLF	80.0	
		Fused	**87.8**	CGLF+PHOG	84.3	83.7

B. MIT Scene Dataset

The MIT Scene dataset [31] 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 and angles, along with a high intra-class variation. Figure 6 (b) shows some sample images from this dataset.

From each class, we use 250 images for training and the rest of the images for testing the performance, and we do this for five random splits. Here YCbCr-HWML is the best single-color descriptor at 88.7% followed closely by oRGB-HWML at 88.2% The combined descriptor FCHWML gives a mean average performance of 90.6%. See Figure 8 for details. Table III shows the category wise descriptor performance on the MIT scene dataset. Table IV compares our result with that of other methods. Note that we tested our descriptor in the rgb color space as well for this dataset and the fused descriptor contains RGB, HSV, rgb, YCbCr and oRGB-HWML descriptors.

IV. Conclusion

We have proposed a new LBP-based texture feature extraction method for color images and combined it with Haar wavelet features and HOG features to generate three new color descriptors: the oRGB-HWML descriptor, the YCbCr-HWML descriptor and the FCHWML descriptor for scene image and object image classification. Results of the experiments using two grand challenge datasets show that our oRGB-HWML and YCbCr-HWML descriptors improve recognition performance over conventional color LBP descriptors. The fusion of multiple color HWML descriptors (FCHWML) shows improvement in the classification performance, which indicates that various color HWML descriptors are not redundant for image classification tasks.

References

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

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

[3] A. Verma, S. Banerji, and C. Liu, "A new color SIFT descriptor and methods for image category classification," in *International Congress on Computer Applications and Computational Science*, Singapore, December 4-6 2010, pp. 819–822.

[4] S. Banerji, A. Verma, and C. Liu, "Novel color LBP descriptors for scene and image texture classification," in *15th Intl. Conf. on Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, Nevada, July 18-21 2011.

[5] W. Niblack, R. Barber, and W. Equitz, "The QBIC project: Querying images by content using color, texture and shape," in *SPIE Conference on Geometric Methods in Computer Vision II*, 1993, pp. 173–187.

[6] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 637–646, 1998.

[7] B. Schiele and J. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *Int. Journal of Computer Vision*, vol. 36, no. 1, pp. 31–50, 2000.

[8] M. Swain and D. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

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

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

[11] A. Bosch, A. Zisserman, and X. Munoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712–727, 2008.

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

[13] 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.

[14] 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.

[15] C. Zhu, C. Bichot, and L. Chen, "Multi-scale color local binary patterns for visual object classes recognition," in *Int. Conf. on Pattern Recognition*, Istanbul, Turkey, August 23-26 2010, pp. 3065–3068.

[16] M. Crosier and L. Griffin, "Texture classification with a dictionary of basic image features," in *Proc. Computer Vision and Pattern Recognition*, Anchorage, Alaska, June 23-28, 2008, pp. 1–7.

[17] 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.

[18] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *ICB'2007*, 2007, pp. 11–18.

[19] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Int. Conf. on Image and Video Retrieval*, Amsterdam, The Netherlands, July 9-11 2007, pp. 401–408.

[20] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes, "Trainable classifier-fusion schemes: An application to pedestrian detection," in *12th International IEEE Conference On Intelligent Transportation Systems*, vol. 1, St. Louis, 2009, pp. 432–437.

[21] 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.

[22] 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.

[23] ——, "Independent component analysis of Gabor features for face recognition," *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 919–928, 2003.

[24] ——, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132–137, 2000.

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

[26] ——, "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.

[27] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 2001.

[28] 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.

[29] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press, 1990.

[30] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Institute of Technology, Tech. Rep. 7694, 2007. [Online]. Available: http://authors.library.caltech.edu/7694

[31] 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.