Object and Scene Image Classification Using Unconventional Color Descriptors

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— *This paper presents novel color, texture and shape descriptors for scene and object image classification and evaluates their performance in unconventional color spaces. First, a new three dimensional Local Binary Pattern (3D-LBP) descriptor is proposed for color and texture feature extraction. Second, a novel color HWML (HOG of Wavelet of Multiplanar LBP) descriptor is derived by computing the histogram of the orientation gradients (HOG) of the Haar wavelet transformation of the original image and the 3D-LBP images. Third, these descriptors are generated in the unconventional color spaces like oRGB, $I_1I_2I_3$, uncorrelated and discriminating color spaces to improve performance over conventional color spaces like RGB and HSV. Fourth, 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 demonstrate the feasibility of the proposed new methods.*

Keywords: The HOG of Wavelet of Multiplanar LBP (HWML) descriptor, the Fused-HWML descriptor, Haar Wavelets, Enhanced Fisher Model (EFM), image search, scene classification

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 significantly larger amount of discriminative information [1], [2], [3]. 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].

[1]Corresponding author

In the past two decades, the recognition and classification of textures using the Local Binary Pattern (LBP) features has been shown to be promising [7], [8], [9]. The Haar wavelet transform has been used for object detection in images and some researchers have combined LBP with Haar-like features for face detection [10]. The Histogram of Orientation Gradients (HOG) descriptor [11], [12] is able to represent an image by storing information about its local shape. Color features when combined with the intensity based texture descriptors are able to outperform many alternatives. 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 multiple color spaces and propose the new oRGB-HWML, YCbCr-HWML and DCS-HWML feature representations, and then integrate them with other color HWML features to produce the novel Fused-HWML descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) [13] and image classification is based on the nearest neighbor classification rule. The effectiveness of the proposed descriptors and classification method is evaluated using two grand challenge datasets: the Caltech 256 image database and the MIT scene database.

2. Related work

In recent years, use of color as a means to image retrieval [14], [3] and object and scene search [4], [1] 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 [15], [16], [17]. 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 [18]. These early methods are sensitive to viewpoint and lighting changes, clutter and occlusions. Hence, global methods were gradually replaced by the part-based methods, which became the popular techniques in the object recognition community.

More recent work on color based image classification appears in [2], [4], [19] 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 [20], and the evaluation of local color invariant descriptors is performed in [5]. The uncorrelated and discriminating colorspaces have been discussed in [19] and the rgb and $I_1I_2I_3$ color spaces have been shown to possess certain advantages over other color spaces in [3]. Fusion of color models, color region detection and color edge detection has been investigated for representation of color images [6]. Some important contributions of color, texture, and shape abstraction for image retrieval have been discussed in Datta et al. [21].

Lately, several methods based on LBP features have been proposed for image representation and classification [8], [9]. 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 [8], [1]. Fusion of different features has been shown to achieve a good image retrieval success rate [1], [9], [22]. Local image descriptors have also been shown to perform well for texture based image retrieval [1], [22].

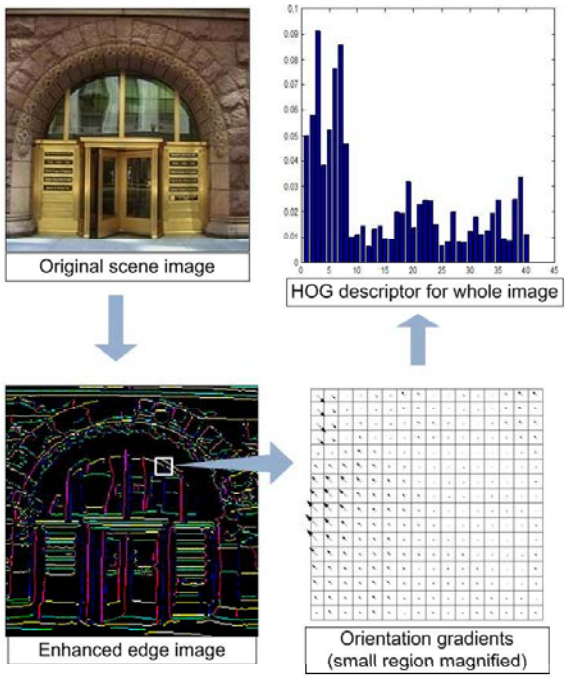

Fig. 1: The Histograms of Orientation Gradients (HOG) descriptor.

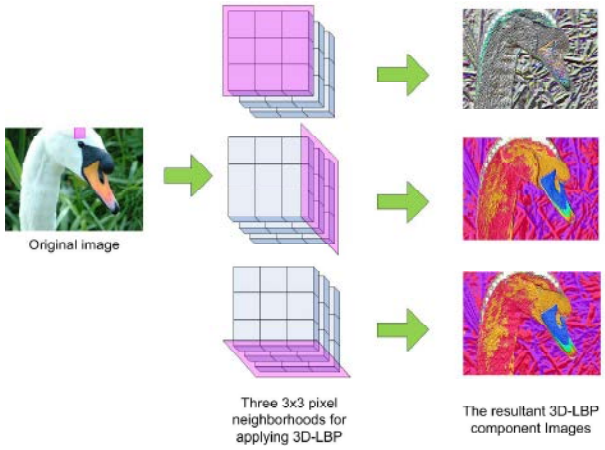

Fig. 2: The proposed 3D-LBP descriptor. A $3 \times 3 \times 3$ pixel region of the original image is magnified to show the 3D-LBP neighborhoods and the resulting LBP images.

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 [10]. The Histogram of Orientation Gradients (HOG) descriptor [11], [12] is able to represent an image by its local shape which is captured by the distribution of 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 [23], [24], [25]. Low-dimensional representation is also important when one considers the computational aspect. PCA has been widely used to perform dimensionality reduction

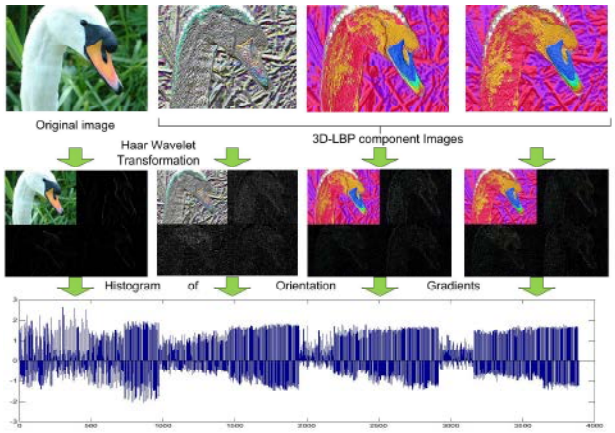

Fig. 3: The proposed HWML descriptor. The original and 3D-LBP images undergo Haar Wavelet Transformation and then HOG is generated for each component of the resulting image and concatenated.

for image indexing and retrieval [13], [26]. The EFM feature extraction method has achieved good success for the task of image representation and retrieval [27].

3. Implementation details

We first review in this section eight color spaces in which our new descriptor is defined, and then discuss the 3D-LBP descriptor which is an improvement upon the traditional LBP descriptor. Next we present the new HWML descriptor in different color spaces and a combined color Fused-HWML descriptor.

3.1 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 [28]. 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 RGB, HSV and YCbCr color spaces have been used by the image processing community for several decades and may be considered fairly conventional. Among these, YCbCr performs very well on scene image search and classification tasks [1]. More recently, several new color spaces have been developed that are less commonly used but they outperform many of the conventional color spaces in scene image classification. Once such color space, the oRGB color space [29] 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 ranges between $[0,1]$:

$$\begin{bmatrix} L \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

Another approach to stabilize RGB images is to decorrelate the RGB components. The $I_1I_2I_3$ color space proposed by Ohta et al. [30] applies a Karhunen Loeve transformation to achieve this. The linear transformation is defined as follows:

$$\begin{aligned} I_1 &= (R+G+B)/3 \\ I_2 &= (R-B)/2 \\ I_3 &= (2G-R-B)/2 \end{aligned} \quad (3)$$

In the color spaces discussed above, the linear transformation matrix is independent of the image content. However, in the next color space that we discuss, different images undergo different transformations based on their content. The Uncorrelated Color Space (UCS), is derived from the RGB color space using a decorrelation method, such as PCA [31]. In the RGB color space, a color image with a spatial resolution of $m \times n$ contains three color component images R, G, and B with the same resolution. Each pixel (x,y) of the color image thus contains three elements corresponding to the red, green, and blue values from the R, G, and B component images which are correlated as well. The UCS decorrelates its component images via a linear transformation $W_U \in \mathbb{R}^{3\times3}$ from the RGB color space

$$\begin{bmatrix} U_1(x,y) \\ U_2(x,y) \\ U_3(x,y) \end{bmatrix} = W_U \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} \quad (4)$$

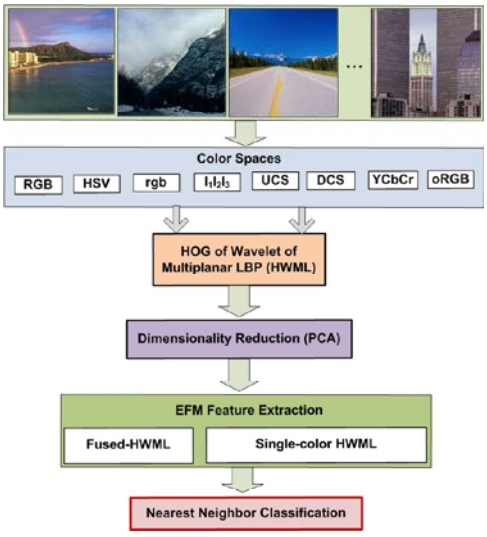

Fig. 4: An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

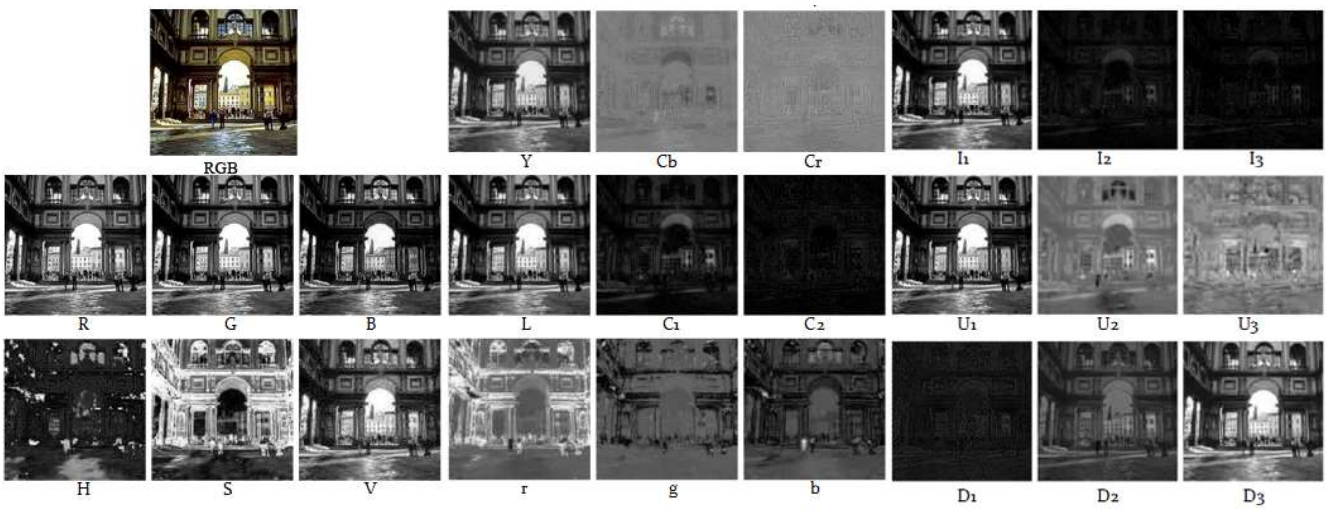

Fig. 5: A sample image from the MIT Scene dataset (top left) is shown split up into various color components.

where $U_1(x,y), U_2(x,y)$, and $U_3(x,y)$ are the values of the uncorrelated component images U_1, U_2, and U_3 in the UCS, $x = 1,2,\ldots,m$ and $y = 1,2,\ldots,n$. The transformation matrix may be derived using PCA. Let $\mathscr{X}$ be the 3-D vector in the RGB color space

$$\mathscr{X} = \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} \tag{5}$$

The covariance matrix of $\mathscr{X}$ can be factorized in the following form:

$$\Sigma_{\mathscr{X}} = W_U^t \Lambda W_U \tag{6}$$

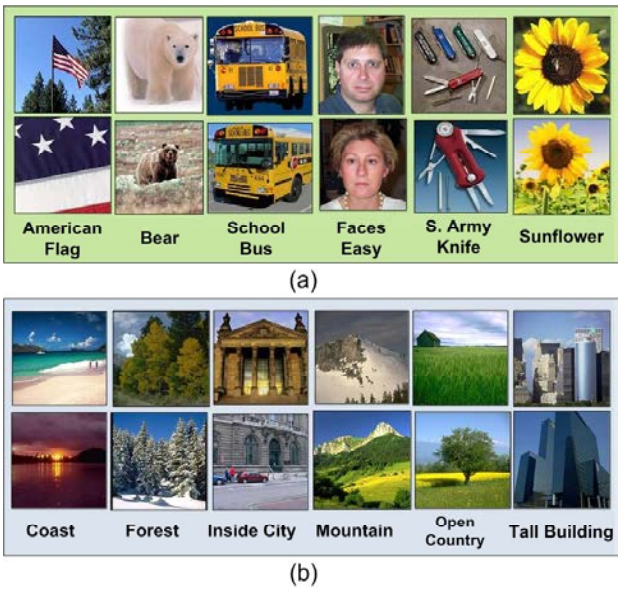

Fig. 6: Some sample images from the (a) Caltech 256 dataset and (b) the MIT Scene dataset.

where $W_U^t \in \mathbb{R}^{3\times 3}$ is an orthonormal eigenvector matrix of the covariance matrix of the random vector $\mathscr{X}$, and $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order. The value of N here is 3.

The Discriminating Color Space (DCS), is derived from the RGB color space by means of discriminant analysis [31]. The (DCS) defines discriminating component images via a linear transformation $W_D \in \mathbb{R}^{3\times 3}$ from the RGB color space

$$\begin{bmatrix} D_1(x,y) \\ D_2(x,y) \\ D_3(x,y) \end{bmatrix} = W_D \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} \tag{7}$$

where $D_1(x,y), D_2(x,y)$, and $D_3(x,y)$ are the values of the discriminating component images D_1, D_2, and D_3 in the DCS, $x = 1,2,\ldots,m$ and $y = 1,2,\ldots,n$. The transformation matrix $W_D \in \mathbb{R}^{3\times 3}$ may be derived through a procedure of discriminant analysis [31]. Let S_w and S_b be the within-class and the between class scatter matrices of the 3-D pattern vector $\mathscr{X}$ respectively. $S_w, S_b \in \mathbb{R}^{3x3}$. The discriminant analysis procedure derives a projection matrix W_D by maximizing the criterion $J_1 = tr(S_w^{-1} S_b)$ [31]. This criterion is maximized when W_D^t consists of the eigenvectors of the matrix $S_w^{-1} S_b$ [31]

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

where W_D^t, Δ are the eigenvector and eigenvalue matrices of $S_w^{-1} S_b$, respectively. Obviously, the transformation matrix W_D depends on the contents of the training image set. The eight colorspaces used by us in this paper and their component images are shown in Figure 5.

3.2 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

Table 1: Category wise descriptor performance (%) on the top 15 classes of the Caltech 256 dataset. Note that the categories are sorted on the Fused-HWML results

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

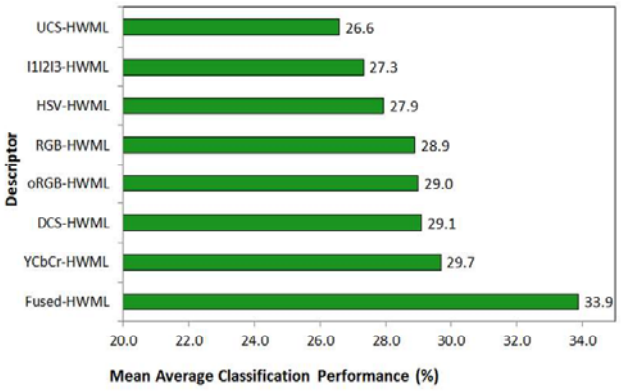

Fig. 7: The mean average classification performance of the proposed HWML descriptor in the RGB, HSV, YCbCr, oRGB, $I_1I_2I_3$, UCS, DCS and fused color spaces using the EFM-NN classifier on the Caltech 256 dataset.

neighboring pixels. Since a color image is represented by a three dimensional matrix, we extended 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 2. 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 3.

3.3 The EFM-NN Classifier

We perform learning and classification using Enhanced Fisher Linear Discriminant Model (EFM) [13]. The EFM method first applies Principal Component Analysis (PCA) to reduce the dimensionality of the input pattern vector. A popular classification method that achieves high separability among the different pattern classes is the Fisher Linear Discriminant (FLD) method. The FLD method, if implemented in an inappropriate PCA space, may lead to overfitting. The EFM method, which applies an eigenvalue spectrum analysis criterion to choose the number of principal components to avoid overfitting, thus improves the generalization performance of the FLD. The EFM method thus derives an appropriate low dimensional representation from the 3DLH descriptor and further extracts the EFM features for pattern classification. We compute similarity score between a training feature vector and a test feature vector using the cosine similarity measure and the nearest neighbor classification rule. Figure 4 gives an overview of multiple feature fusion methodology, the EFM feature extraction method, and the classification stages.

Table 2: Comparison of the Classification Performance (%) with Other Methods on Caltech 256 Dataset

#train	#test	HWML		[4]	
15360	5120	DCS	**30.5**	-	
		oRGB	**30.7**		
		YCbCr	**31.1**		
		Fused	**35.6**		
12800	6400	oRGB	**29.0**	oRGB-SIFT	23.9
		DCS	**29.1**	CSF	30.1
		YCbCr	**29.7**	CGSF	**35.6**
		Fused	33.9		

4. Experimental results

4.1 Caltech 256 Dataset

The Caltech 256 dataset [33] 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, a maximum of 827 images and 119 mean number of images per category. 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. A few sample images from this dataset can be seen in Figure 6(a).

On this dataset, we conduct experiments for HWML descriptors from seven 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 [33]. Figure 7 shows

Table 3: Category wise descriptor performance (%) on the MIT Scene dataset. Note that the categories are sorted on the Fused-HWML results

Category	Fusion	YCbCr	DCS	oRGB	RGB	HSV	UCS	$I_1I_2I_3$	rgb
forest	97	**98**	97	96	97	97	96	97	95
coast	**94**	90	91	92	93	90	90	91	91
street	94	88	89	**95**	89	89	86	90	88
inside city	92	92	**93**	92	92	88	89	87	89
mountain	**92**	90	89	90	89	88	88	87	86
tall building	**92**	89	87	86	87	88	85	85	86
highway	**90**	88	88	88	86	86	82	88	84
open country	**79**	75	72	76	71	74	70	72	68
Mean	**91.3**	**88.7**	**88.3**	**88.2**	**88.1**	**87.5**	**87.1**	**85.9**	**85.7**

the detailed performance of our EFM-NN classification technique on this dataset. The best recognition rate that we obtain is 33.9%, which is a very respectable value for a dataset of this size and complexity. It can be seen from our previous work [4] 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, processor-intensive SIFT based methods achieve 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 4%. 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 2 compares our results with those of SIFT-based methods. Table 1 shows the descriptor performance for the top 15 categories from this dataset. The Fused-HWML recognition rates for the top 15 categories lie between 76% and 100% with five categories having full success rate.

Table 4: Comparison of the Classification Performance (%) with Other Methods on the MIT Scene Dataset

#train	#test	HWML		[1]		[32]
		oRGB	**88.2**			
		DCS	**88.3**	CLF	86.4	-
2000	688	YCbCr	**88.7**	CGLF	86.6	
		Fused	**91.3**	CGLF+PHOG	89.5	
		DCS	**85.6**			
		YCbCr	**85.7**	CLF	79.3	
800	1888	oRGB	**85.8**	CGLF	80.0	
		Fused	**87.8**	CGLF+PHOG	84.3	83.7

4.2 MIT Scene Dataset

The MIT Scene dataset [32] 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. Some images from this dataset can be seen in Figure 6(b).

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 DCS-HWML at 88.3%. The combined descriptor Fused-HWML gives a mean average performance of 91.3%. See Figure 8 for details. Table 3 shows the category wise descriptor performance on the MIT scene dataset. We achieve the best classification performance for categories like forest and coast where there are disctinctive visual elements like trees and the ocean. The worst classification performance is seen

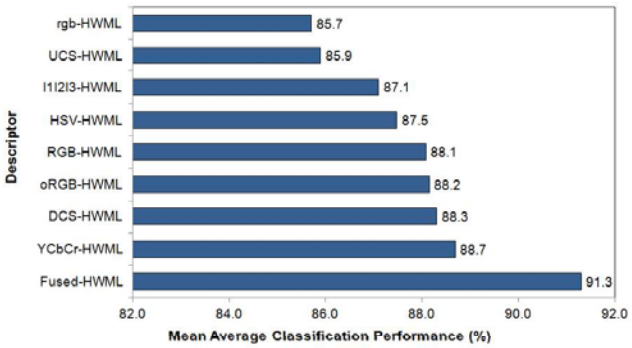

Fig. 8: The mean average classification performance of the proposed HWML descriptor in the RGB, rgb, HSV, YCbCr, oRGB, $I_1I_2I_3$, UCS, DCS and fused color spaces using the EFM-NN classifier on the MIT scene dataset.

for the open country category where there are very few distinguishing visual features. For the other seven categories, the Fused-HWML descriptor classifies with 90% or above success rate. Table 4 compares our result with that obtained by 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, UCS, DCS, $I_1I_2I_3$ and oRGB-HWML descriptors.

5. 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 several new color descriptors: the oRGB-HWML descriptor, the YCbCr-HWML descriptor, the DCS-HWML descriptor and the Fused-HWML descriptor for scene image and object image classification. Results of the experiments using two grand challenge datasets show that our oRGB-HWML, DCS-HWML and YCbCr-HWML descriptors improve recognition performance over conventional color LBP descriptors. The fusion of multiple color HWML descriptors (Fused-HWML) shows significant improvement in the classification performance, which indicates that various color HWML descriptors are not redundant for scene image classification tasks.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[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. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Int. Conf. on Pattern Recognition*, Jerusalem, Israel, 1994, pp. 582–585.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[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, "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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

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

[26] 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.

[27] ——, "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.

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

[29] 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.

[30] Y. Ohta, *Knowledge-Based Interpretation of Outdoor Natural Color Scenes*. Pitman Publishing, London, 1985.

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

[32] 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.

[33] 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