Novel Color Gabor-LBP-PHOG (GLP) Descriptors for Object and Scene Image Classification

Atreyee Sinha*
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA
as739@njit.edu

Sugata Banerji
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA
sb256@njit.edu

Chengjun Liu
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA
chengjun.liu@njit.edu

ABSTRACT

This paper presents a novel set of color descriptors for object and scene image classification. We first introduce a new Gabor-PHOG (GPHOG) descriptor by concatenating the Pyramid of Histograms of Oriented Gradients (PHOG) of the local Gabor filtered images. Second, we derive the Gabor-LBP (GLBP) descriptor by accumulating the Local Binary Patterns (LBP) histograms of all the component images produced by applying Gabor filters. Then, by combining the GPHOG and the GLBP descriptors using an optimal feature representation method, we propose a novel Gabor-LBP-PHOG (GLP) image descriptor which performs well on different image categories. Next, we make a comparative assessment of the classification performance of the GLP descriptor in six different color spaces. Finally, we present a novel Fused Color GLP (FC-GLP) feature by integrating the PCA features of the six color GLP descriptors. The Principal Component Analysis (PCA) and the Enhanced Fisher Model (EFM) are applied for feature extraction and the nearest neighbor classification rule is used for classification. The effectiveness of the proposed GLP and FC-GLP feature vectors for image classification is evaluated using three grand challenge datasets, namely the Caltech 256 dataset, the MIT Scene dataset and the UIUC Sports Event dataset.

Keywords

Gabor-PHOG (GPHOG), Gabor-LBP (GLBP), Gabor-LBP-PHOG (GLP), Fused Color GLP (FC-GLP), Enhanced Fisher Model (EFM), color spaces, image search

1. INTRODUCTION

Object and scene recognition is an important part of content based image classification and retrieval. With high variation in pose, angle and illumination, object and scene classification is a very challenging task. A good classification framework should address the key issues of discriminatory feature extraction as well as efficient and accurate classification. Color provides powerful discriminating information as humans can distinguish thousands of colors, compared to about only two dozen shades of gray [14], and color based image search can be very effective for image classification tasks [1], [34], [38]. Some desirable properties of the descriptors defined in different color spaces include relative stability over changes in photographic conditions such as varying illumination. Global color features such as the color histograms and local invariant features provide varying degrees of success against image variations such as rotation, viewpoint and lighting changes, clutter and occlusions [8]. Shape, texture and local features also provide important cues for content based image classification and retrieval. Recent works employing local texture features such as Local Binary Patterns (LBP) [32] have shown promising results for texture and scene recognition [1]. Shape is another important feature which can be useful for understanding an image. Local object shape and the spatial layout of the shape within an image can be described by the Pyramid of Histograms of Oriented Gradients (PHOG) descriptor [6].

In this paper, we attempt to develop novel color descriptors exploring the concept of how people understand and recognize object and scene images. We subject the image to a series of Gabor wavelet transformations, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells [30]. The novelty of this paper is in the construction of several feature vectors based on Gabor filters. Specifically, we first present a novel Gabor-PHOG (GPHOG) descriptor by concatenating the Pyramid of Histograms of Oriented Gradients (PHOG) of the components of the images produced by the result of applying a series of Gabor filters. Second, we fuse the GPHOG with the Gabor-LBP (GLBP) descriptor, formed by taking the LBP histograms of the Gabor filtered images in different orientations to design the robust Gabor-LBP-PHOG (GLP) feature vector. We then assess our GLP feature vector in six different color spaces, and propose several new color GLP feature representations. We further extend this concept by integrating the PCA features of the six color GLP vectors to produce the novel Fused Color GLP (FC-GLP) descriptor. Feature extraction applies the Principal Component Analysis (PCA) and the Enhanced Fisher Model (EFM) [26], and

*Corresponding author

ICVGIP '12, December 16-19, 2012, Mumbai, India

image classification is based on the nearest neighbor classification rule. Finally, the robustness of the proposed descriptors and classification method is evaluated using three datasets: the Caltech 256 grand challenge image dataset, the UIUC Sports Event dataset and the MIT Scene dataset.

2. BACKGROUND

In the early days of image search, Swain and Ballard designed a color indexing system which used color histogram for image inquiry from a large image database [36]. Notable contributions on color based image classification appear in [38], [24], [23] that propose several new color spaces and methods for face, object and scene image classification. It has been shown that fusion of color features achieve higher classification performance in the works of [1], [39], [35]. Datta et al. [10] discussed the importance of color, texture and shape abstraction for content based image retrieval.

Several researchers have described the biological relevance and computational properties of Gabor wavelets for image analysis [30], [11]. Donato et al. [12] showed experimentally that the Gabor wavelet representation is optimal for classifying facial actions. In texture recognition, a Gabor filter-based approach has been successfully used [29]. Some researchers have also employed the LBP histogram sequences of the Gabor wavelets for face image recognition [18], [40].

A robust feature extraction method that has the ability to learn meaningful low-dimensional patterns in spaces of very high dimensionality [21], [25], [28] is needed for efficient retrieval. Low-dimensional representation is also important when one considers the computational aspect. PCA has been widely used to perform dimensionality reduction for image indexing and retrieval [26], [22]. The EFM feature extraction method has achieved good success for the task of image classification and retrieval [27].

3. IMPLEMENTATION DETAILS

In this section, we will first briefly review the color spaces in which our new descriptors are defined, and then discuss the proposed novel descriptors and methodology for image classification.

3.1 Color spaces

A color image is defined by a function of two spatial variables and one spectral variable, where the spectral dimension is usually sampled to the red (R), green (G), and blue (B) spectral bends, known as the primary colors. The commonly used color space is the RGB color space. Other color spaces are usually derived from the RGB color space by means of either linear or nonlinear transformations. 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 [14].

The YIQ color space is adopted by the NTSC (National Television System Committee) video standard in reference to RGB NTSC. The I and Q components are derived from the U and V counterparts of the YUV color space via a clockwise rotation (33°) [34], and is defined as follows:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5957 & -0.2745 & -0.3213 \\ 0.2115 & -0.5226 & 0.3111 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

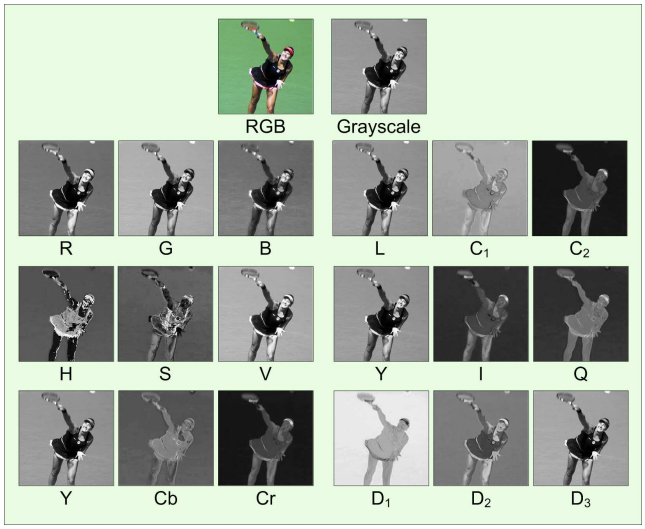

Figure 1: A sample color image (labeled RGB) is shown split up into various color components of the RGB, oRGB, HSV, YIQ, YCbCr and DCS as well as grayscale.

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.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

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

The oRGB color space [7] 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 $[0, 1]$.

$$\begin{bmatrix} L \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

The Discriminating Color Space (DCS), is derived from the RGB color space by means of discriminant analysis [13]. 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. Let $\mathcal{X}$ be the 3-D vector in the RGB color space

$$\mathcal{X} = \begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} \quad (4)$$

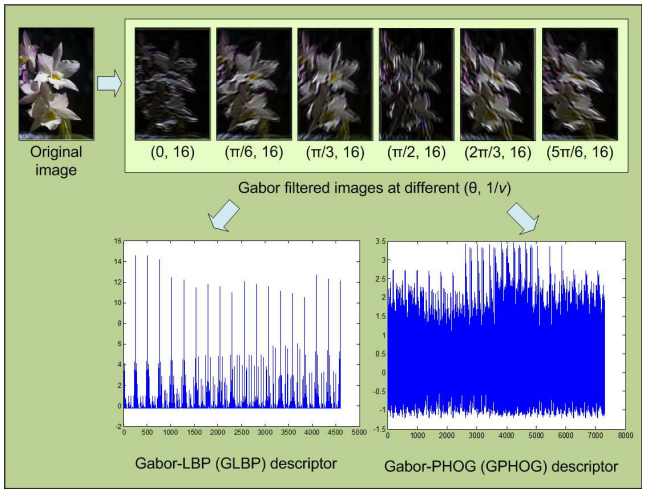

Figure 2: The generation of the Gabor-LBP (GLBP) and the Gabor-PHOG (GPHOG) descriptors.

The DCS [24] defines discriminating component images via a linear transformation $W_D \in \mathbb{R}^{3\times3}$ from the RGB color space. The linear transformation is defined as

$$\begin{bmatrix} D_1(x,y) \\ D_2(x,y) \\ D_3(x,y) \end{bmatrix} = W_D \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} \quad (5)$$

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 $W_D \in \mathbb{R}^{3\times3}$ may be derived through a procedure of discriminant analysis [13]. Let S_w and S_b be the within-class and the between class scatter matrices of the 3-D pattern vector $\mathcal{X}$ respectively. $S_w, S_b \in \mathbb{R}^{3x3}$. The discriminant analysis procedure derives a projection matrix W_D by maximizing the criterion $J_1 = tr(S_w^{-1}S_b)$ [13]. This criterion is maximized when W_D^t consists of the eigenvectors of the matrix $S_w^{-1}S_b$ [13]. Figure 1 shows the grayscale and the color components of a sample image in the six color spaces used by us in this paper.

3.2 Novel Color GLP Descriptors for Object and Scene Image Classification

The Gabor filter, which is based on the mathematical model of filter theory, is considered to be a good model for human visual receptive fields and hence the Gabor filter-based approach has been used in this paper for subsequent extraction of novel feature vectors.

3.2.1 The Gabor-PHOG (GPHOG) and the Gabor-LBP (GLBP) descriptors

A Gabor filter is obtained by modulating a sinusoid with a Gaussian distribution. In the case of 2D signals such as images, a Gabor filter is defined as:

$$g_{\nu,\theta,\phi,\sigma,\gamma}(x,y) = \exp(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2})\exp(i(2\pi\nu x'+\phi)) \quad (6)$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, and ν, θ, ϕ, σ, γ denote the spatial frequency of the sinusoidal factor, orientation of the normal to the parallel stripes of a Gabor function, phase offset, standard deviation of the Gaussian kernel and the spatial aspect ratio specifying the ellipticity of the support of the Gabor function respectively. For a grayscale image $f(x,y)$, the Gabor filtered image is produced by convolving the input image with the real and imaginary components of a Gabor filter [18].

The Pyramid of Histograms of Oriented Gradients (PHOG) [6] descriptor, inspired from the Histograms of Oriented Gradients (HOG) [9] and the image pyramid representation of Lazebnik et al. [17], represents local image shape and its spatial layout, together with a spatial pyramid kernel. 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 [6].

The Local Binary Patterns (LBP) method derives the texture description of a grayscale image by comparing a center pixel with its neighbors [32]. In particular, for a 3×3 neighborhood of a pixel $\mathbf{p} = [x,y]^t$, $\mathbf{p}$ is the center pixel is used as a threshold, and the neighbors of the pixel $\mathbf{p}$ are defined as $N(\mathbf{p},i) = [x_i,y_i]^t$, $i = 0,1,\cdots,7$, where i is the number used to label the neighbor.

We used the Gabor wavelet representation for subsequent extraction of our feature vectors as it captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity. We subject each of the three color components of the image to different even symmetric Gabor filters [2], [16]. For all our experiments, we choose the parameter values as $\phi = 0$, $\sigma = 8$, $\gamma = 1$, $\nu = 1/16$ and $\theta = [0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6]$, and the size of

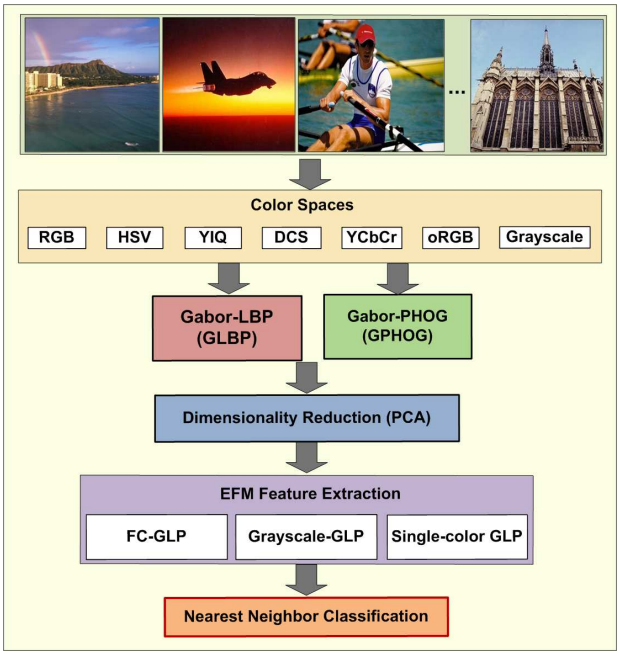

Figure 3: An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

Figure 4: Some sample images from the Caltech 256 dataset.

the filters used is 33×33. We derive the novel Gabor-PHOG (GPHOG) feature vector by concatenating the PHOG of the components of the Gabor filtered images. It should be noted that we computed the PHOG with two levels for generating our GPHOG descriptor. LBP histograms of each component of the images produced as a result of applying Gabor filters in one scale and six orientations as stated earlier are calculated and concatenated to form the Gabor-LBP (GLBP) descriptor. Figure 2 shows the generation of the GLBP and GPHOG features.

3.2.2 The GLP and the FC-GLP descriptors

We pass both the GPHOG and the GLBP feature vectors through a dimensionality reduction phase using a popular method such as Principal Component Analysis (PCA) to derive the most expressive and meaningful features. The features so produced after applying the optimal feature representation technique on the two descriptors are normalized to zero mean and unit standard deviation and then integrated to develop the new GLP feature vector, the performance of which is measured in six different color spaces, namely RGB, HSV, oRGB, YCbCr, YIQ and DCS as well as in grayscale. We use PCA for the optimal representation of our color GLP vectors with respect to minimum mean square error, and the PCA features of the six normalized color GLP descriptors are further combined to form the novel Fused Color GLP (FC-GLP) descriptor which outperforms the classification results of the individual color GLP features.

3.3 The EFM-NN Classifier

We perform learning and classification using Enhanced Fisher Linear Discriminant Model (EFM) [26] and the nearest neighbor classification rule. 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 applies an eigenvalue spectrum analysis criterion to choose the number of principal components to avoid overfitting and improves the generalization performance of the FLD. The EFM method thus derives an appropriate low dimensional representation from the GLP 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 3 gives an overview of multiple feature fusion methodology, the EFM feature extraction method, and the classification stages.

4. EXPERIMENTS

In this section, we will first give a brief description of the datasets used for our experiments, and then discuss the classification performance of our novel color GLP and FC-GLP descriptors.

4.1 Datasets

We tested our descriptors using three popular and publicly available datasets, namely: the Caltech 256 dataset, the UIUC Sports Event dataset, and the MIT Scene dataset.

4.1.1 The Caltech 256 Dataset

The Caltech 256 dataset [15] 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 and at most 827 images. The mean number of images per category is 119. 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.

Figure 5: Some sample images from (a) the UIUC Sports Event dataset, (b) the MIT Scene dataset.

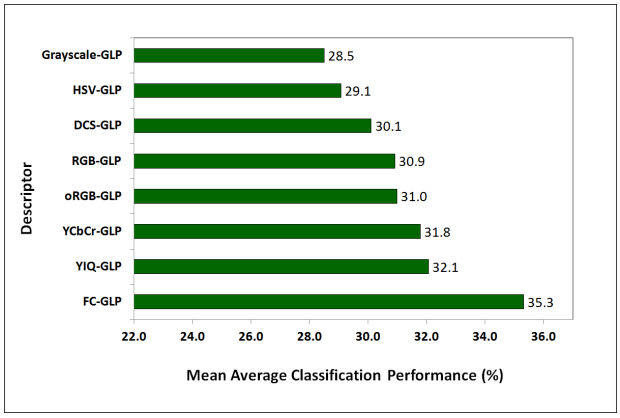

Figure 6: The mean average classification performance of the proposed color GLP and FC-GLP descriptors on the Caltech 256 dataset.

The average size of each image is 351×351 pixels. Figure 4 shows some sample images from this dataset. 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 [15].

4.1.2 *The UIUC Sports Event Dataset*

The UIUC Sports Event dataset [19] contains 8 sports event categories: rowing (250 images), badminton (200 images), polo (182 images), bocce (137 images), snowboarding (190 images), croquet (236 images), sailing (190 images), and rock climbing (194 images). A few sample images of this dataset can be seen in figure 5(a).

From each class, we use 70 images for training and 60 images for testing the classification performance of our descriptors, and we do this for five random splits. Other researchers [3], [20] have also reported using the same number of images for training and testing.

4.1.3 *The MIT Scene Dataset*

The MIT Scene dataset [33] has 2,688 images classified as eight categories: 360 coast, 328 forest, 260 highway, 308 inside of cities, 374 mountain, 410 open country, 292 streets, and 356 tall buildings. Some sample images from this dataset are shown in figure 5(b). All of the images are in color, in JPEG format, and the size of each image is 256×256 pixels. There is a large variation in light and angles along with a high intra-class variation. From each class, we use 250 images for training and the rest of the images for testing the performance. We also perform a second set of experiments for this dataset using 100 training images from each class and the rest of the images for testing. For each of the experiments, we do a five-fold cross validation.

Table 1: Comparison of the Classification Performance (%) with Other Methods on Caltech 256 Dataset.

Descriptor		Performance (%)
#train = 12800, #test = 6400		
oRGB-SIFT	[38]	23.9
gray-PHOW		25.9
color-PHOW		29.9
CSF	[38]	30.1
FC-GLP	**(Ours)**	**35.3**
CGSF	[38]	35.6

Table 2: Comparison of the Classification Performance (%) with Other Methods on the UIUC Sports Event Dataset.

Descriptor		Performance (%)
#train = 560, #test = 480		
SIFT+GGM	[19]	73.4
OB	[20]	76.3
gray-PHOW		76.4
CA-TM	[31]	78.0
color-PHOW		79.0
SIFT+SC	[3]	82.7
FC-GLP	**(Ours)**	**84.3**
HMP	[3]	85.7

4.2 Results and Discussion

In this section, we evaluate the performance of our proposed GLP and FC-GLP descriptors in the three datasets, and also compare it with some popular descriptors. Specifically, we compare our FC-GLP descriptor with the popular SIFT-based Pyramid Histograms of visual Words (PHOW) descriptor [5] on all three datasets. To compare the proposed FC-GLP descriptor with the popular SIFT-based feature, we generate the Pyramid Histograms of visual Words (PHOW) feature vector [5] using the software package VLFeat [37]. To compare the classification performance of our proposed descriptor, we use the gray PHOW as well as the color PHOW feature vectors. For both PHOW and FC-GLP, we use PCA to obtain the most expressive features and the EFM-NN classifier in order to make a fair comparison. In addition,

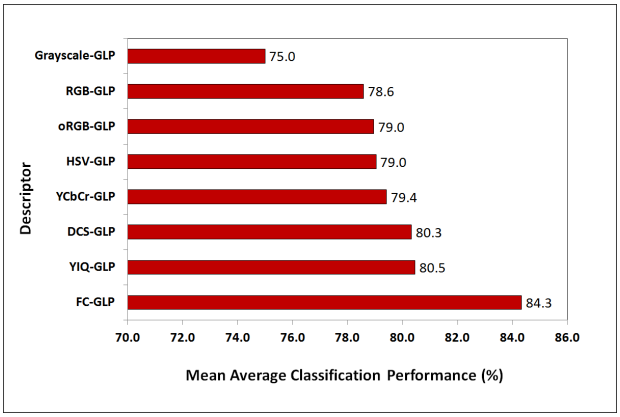

Figure 7: The mean average classification performance of the proposed GLP descriptor in individual color spaces as well as after fusing them on the UIUC Sports Event dataset.

Table 3: Comparison of the Classification Performance (%) with Other Methods on the MIT Scene Dataset.

Descriptor		Performance (%)
#train = 800, #test = 1888		
CLF	[1]	79.3
CGLF	[1]	80.0
gray-PHOW		82.5
SE	[33]	83.7
color-PHOW		84.3
CGLF+PHOG	[1]	84.3
C4CC	[4]	86.7
FC-GLP	**(Ours)**	**87.5**
#train = 2000, #test = 688		
gray-PHOW		86.2
CLF	[1]	86.4
CGLF	[1]	86.6
color-PHOW		89.3
CGLF+PHOG	[1]	89.5
FC-GLP	**(Ours)**	**91.3**

we also compare the classification performance achieved by our FC-GLP descriptor coupled with the EFM-NN classifier to the image classification performance of some other popular methods as reported in published papers.

In the Caltech 256 dataset, YIQ-GLP performs the best among single-color descriptors giving 32.1% success followed by YCbCr-GLP and oRGB-GLP with 31.8% and 31.0% classification rates respectively. Figure 6 shows the success rates of the GLP descriptors for this dataset. The FC-GLP descriptor here achieves a success rate of 35.3%. Table 1 compares our results with other methods.

In the UIUC Sports Event dataset, the YIQ-GLP is the best single-color descriptor at 80.5% followed by DCS-GLP and YCbCr-GLP respectively. The combined descriptor FC-GLP gives a mean average performance of 84.3%. See Figure 7 for details. Table 2 compares our result with that

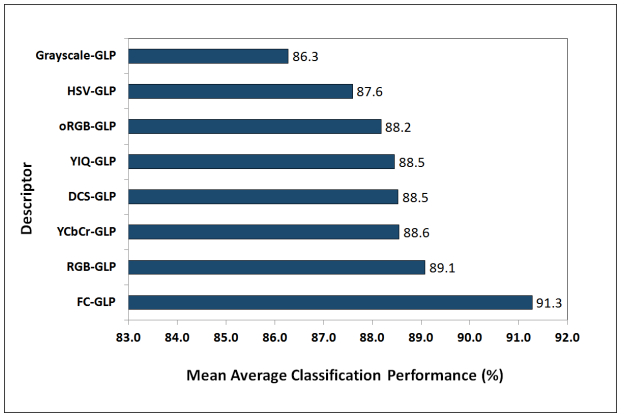

Figure 8: The mean average classification performance of the proposed GLP descriptor in individual color spaces as well as after fusing them on the MIT Scene dataset.

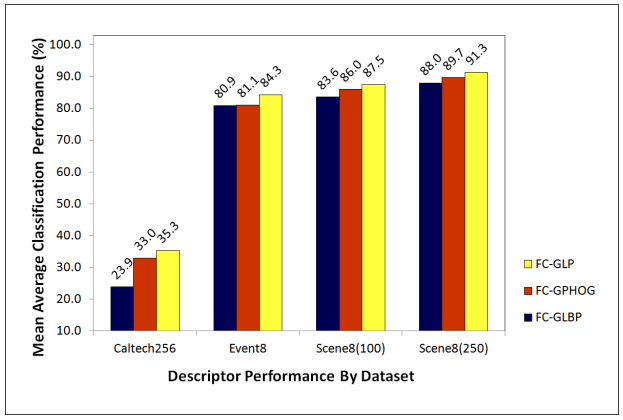

Figure 9: The comparative mean average classification performance of the FC-GLBP, FC-GPHOG and FC-GLP descriptors on the Caltech 256, UIUC Sports Event and MIT Scene (with 100 and 250 training images per class) datasets.

obtained by other researchers. The category wise recognition performance of our GLP descriptors on this dataset is shown in table 4.

For the MIT Scene dataset, using 250 training images per class, the RGB-GLP is the best single-color descriptor at 89.1% followed closely by YCbCr-GLP and DCS-GLP. The combined descriptor FC-GLP gives a mean average performance of 91.3%. See Figure 8 for details. Table 3 compares our result with that of other methods. Table 5 shows the class wise classification rates for this dataset on applying the proposed GLP descriptors.

Figure 9 gives a comparison of the FC-GLBP, FC-GPHOG descriptors and their fusion (FC-GLP) for image classification in the three datasets used for our experiments. It should be noted that the generation time of the GPHOG and the GLBP features varies linearly with the number of pixels in the input image. We observe that the six color GLP features beat the recognition performance of the Grayscale-GLP descriptor which show information contained in color images can be significantly more useful than that in grayscale images for classification. Furthermore, the fusion of multiple color GLP descriptors (FC-GLP) achieves significant increase in the classification performance over individual color GLP descriptors, which implies that various color GLP descriptors are not completely redundant for image classification tasks.

5. CONCLUSION

We have presented a new Gabor-based local, texture, shape and color feature extraction method inspired by PHOG and fused it with GLBP features using an optimal feature representation method such as PCA, to propose the robust GLP descriptor and measure its performance in six different color spaces as well as in grayscale. We then fuse the six color GLP descriptors as a feature set to further develop the novel FC-GLP image descriptor which exceeds or achieves comparable performance to some of the best classification performances reported elsewhere. Experimental results carried out using three grand challenge datasets show that our FC-GLP de-

Table 4: Category wise GLP descriptor performance (%) on the UIUC Sports Event dataset. Note that the categories are sorted on the FC-GLP results

Category	FC	YIQ	DCS	YCbCr	HSV	oRGB	RGB	Grayscale
rock climbing	**96**	94	94	93	93	94	93	90
sailing	**94**	**94**	**94**	**94**	93	92	**94**	92
badminton	**93**	88	86	88	85	87	87	87
rowing	**88**	87	86	85	84	86	85	82
snow boarding	**88**	84	83	83	83	82	81	75
polo	**86**	76	76	78	81	74	78	72
croquet	**75**	74	71	69	65	67	68	60
bocce	**55**	47	53	46	48	50	43	42
Mean	**84.3**	**80.5**	**80.3**	**79.4**	**79.0**	**79.0**	**78.6**	**75.0**

Table 5: Category wise GLP descriptor performance (%) on the MIT Scene dataset. Note that the categories are sorted on the FC-GLP results

Category	FC	RGB	YCbCr	DCS	YIQ	oRGB	HSV	Grayscale
forest	**97**	**97**	96	96	96	96	**97**	96
highway	**94**	90	88	90	88	90	90	88
tall building	94	**95**	93	94	94	92	93	94
street	93	92	90	**94**	92	90	90	89
coast	**93**	89	**93**	88	91	91	90	87
mountain	**91**	87	87	88	89	86	87	72
inside city	88	**90**	86	87	84	87	83	86
open country	**80**	77	76	72	73	74	69	69
Mean	**91.3**	**89.1**	**88.6**	**88.5**	**88.5**	**88.2**	**87.6**	**86.3**

scriptor improves classification performance over the GLBP and GPHOG descriptors and can be successfully applied for object and scene image classification.

6. REFERENCES

[1] S. Banerji, A. Verma, and C. Liu. Novel color LBP descriptors for scene and image texture classification. In *15th International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pages 537–543, Las Vegas, Nevada, July 18-21 2011.

[2] T. Barbu. Novel automatic video cut detection technique using gabor filtering. *Computers and Electrical Engineering*, 35(5):712–721, September 2009.

[3] L. Bo, X. Ren, and D. Fox. Hierarchical matching pursuit for image classification: Architecture and fast algorithms. In *Advances in Neural Information Processing Systems*, December 2011.

[4] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via pLSA. In *Proceedings of the European Conference on Computer Vision*, 2006.

[5] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In *Proceedings of the 11th International Conference on Computer Vision, Rio de Janeiro, Brazil*, 2007.

[6] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *International Conference on Image and Video Retrieval*, pages 401–408, Amsterdam, The Netherlands, July 9-11 2007.

[7] M. Bratkova, S. Boulos, and P. Shirley. oRGB: A practical opponent color space for computer graphics. *IEEE Computer Graphics and Applications*, 29(1):42–55, 2009.

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

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, pages 886–893, Washington, DC, USA, 2005.

[10] R. Datta, D. Joshi, J. Li, and J. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):509–522, 2008.

[11] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856, 1980.

[12] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

[13] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1990.

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

[15] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007.

[16] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti. Filterbank-based fingerprint matching. *IEEE Transactions on Image Processing*, 9(5):846–859, 2000.

[17] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, Washington, DC, USA, 2006.

[18] H. Lee, Y. Chung, J. Kim, and D. Park. Face image retrieval using sparse representation classifier with gabor-lbp histogram. In *WISA*, pages 273–280, Berlin, Heidelberg, 2010.

[19] L.-J. Li and L. Fei-Fei. What, where and who? classifying event by scene and object recognition. In *Proceedings of IEEE International Conference in Computer Vision.*, 2007.

[20] L.-J. Li, H. Su, E. P. Xing, and L. Fei-Fei. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Neural Information Processing Systems*, Vancouver, Canada, December 2010.

[21] C. Liu. A Bayesian discriminating features method for face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):725–740, 2003.

[22] C. Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):572–581, 2004.

[23] C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):725–737, 2006.

[24] C. Liu. Learning the uncorrelated, independent, and discriminating color spaces for face recognition. *IEEE Transactions on Information Forensics and Security*, 3(2):213–222, 2008.

[25] C. Liu and H. Wechsler. Evolutionary pursuit and its application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):570–582, 2000.

[26] C. Liu and H. Wechsler. Robust coding schemes for indexing and retrieval from large face databases. *IEEE Transactions on Image Processing*, 9(1):132–137, 2000.

[27] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[28] C. Liu and H. Wechsler. Independent component analysis of Gabor features for face recognition. *IEEE Transactions on Neural Networks*, 14(4):919–928, 2003.

[29] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[30] S. Marcelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70:1297–1300, 1980.

[31] Z. Niu, G. Hua, X. Gao, and Q. Tian. Context aware topic model for scene recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2743–2750, Providence, RI, USA, June 16-21 2012.

[32] T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In *International Conference on Pattern Recognition*, pages 582–585, Jerusalem, Israel, 1994.

[33] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[34] 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*, 19(7), 2005.

[35] H. Stokman and T. Gevers. Selection and fusion of color models for image feature detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):371–381, 2007.

[36] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[37] A. Vedaldi and B. Fulkerson. Vlfeat – an open and portable library of computer vision algorithms. In *Proceedings of the 18th Annual ACM International Conference on Multimedia*, 2010.

[38] 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*, pages 819–822, Singapore, December 4-6 2010.

[39] A. Verma and C. Liu. Novel EFM-KNN classifier and a new color descriptor for image classification. In *20th IEEE Wireless and Optical Communications Conference (Multimedia Services and Applications)*, Newark, New Jersey, USA, April 15-16 2011.

[40] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. In *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 1*, pages 786–791, Washington, DC, USA, 2005.