Gabor-Based Novel Color Descriptors for Object and Scene Image Classification

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

Abstract— *This paper presents several novel Gabor-based color descriptors for object and scene image classification. Firstly, a new Gabor-HOG descriptor is proposed for image feature extraction. Secondly, the Gabor-LBP descriptor derived by concatenating the Local Binary Patterns (LBP) histograms of all the component images produced by applying Gabor filters is integrated with the Gabor-HOG using an optimal feature representation method to introduce a novel Gabor-LBP-HOG (GLH) image descriptor which performs well on different object and scene image categories. Finally, the Enhanced Fisher Model (EFM) is applied for discriminatory feature extraction and the nearest neighbor classification rule is used for image classification. The robustness of the proposed GLH feature vector is evaluated using three grand challenge datasets, namely the Caltech 256 dataset, the MIT Scene dataset and the UIUC Sports Event dataset.*

Keywords: The Gabor-HOG descriptor, the Gabor-LBP descriptor, the GLH descriptor, EFM, color fusion, color spaces, image search

1. Introduction

Content-based image classification and retrieval is an important research area in image processing and analysis. A good image classification system should consider both extraction of meaningful features from the image as well as accurate classification. Color provides powerful discriminating information as humans can distinguish thousands of colors, compared to about only two dozen shades of gray [1], and color based image search has been found to be very effective for image classification tasks [2], [3], [4]. Color features derived from different color spaces exhibit different properties, most important of which is that they need to be stable under changing viewing conditions, such as changes in illumination, shading and highlights. 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]. Recent works on employing local texture features such as Local Binary Patterns (LBP) [7] have been promising for object and scene recognition [2], [8]. Local object appearance and shape within an image can be described by the Histogram of Oriented Gradients (HOG) that stores distribution of edge orientations within an image [9].

The motivation behind this work lies in the concept of how people understand and recognize objects and scenes. 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 [10]. The Gabor filters exhibit desirable characteristics of spatial locality and orientation selectivity [11], [12]. The novelty of this paper is in the construction of several feature vectors based on Gabor filters. Specifically, we first present a novel Gabor-HOG descriptor by taking the Histogram of Oriented Gradients of the components of the images produced by the result of applying Gabor filters in different scales and orientations. We then fuse the Gabor-HOG with the Gabor-LBP descriptor, formed by taking the LBP histograms of the Gabor filtered images, in the PCA space to design the robust Gabor-LBP-HOG (GLH) feature vector. We further extend this concept to different color spaces and propose several new GLH feature representations, and finally integrate them with other color GLH features to produce the novel Fused Color GLH (FC-GLH) descriptor. Feature extraction applies the Enhanced

Fig. 1: Some sample images from the Caltech 256 dataset.

[1]Corresponding author

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 three datasets: the Caltech 256 grand challenge image dataset, the UIUC Sports Event dataset and the MIT Scene dataset.

2. Related work

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 [14]. More recent work on color based image classification appears in [4], [15] that propose several new color spaces and methods for face, object and scene classification. It has been shown that fusion of color features achieve higher classification performance in the works of [2], [16], [6]. Datta et al. [17] 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 [10], [18]. Lades et al. [19] used Gabor wavelets for face recognition using the Dynamic Link Architecture (DLA) framework. Lately, Donato et al. [20] showed experimentally that the Gabor wavelet representation is optimal for classifying facial actions. Methods based on LBP features have been proposed for image representation and classification [8]. Extraction of LBP features is computationally efficient and with the use of multi-scale filters invariance to scaling and rotation can be achieved [8], [2]. Some researchers have also employed the LBP histogram sequences of the Gabor wavelets for face image recognition [21], [22].

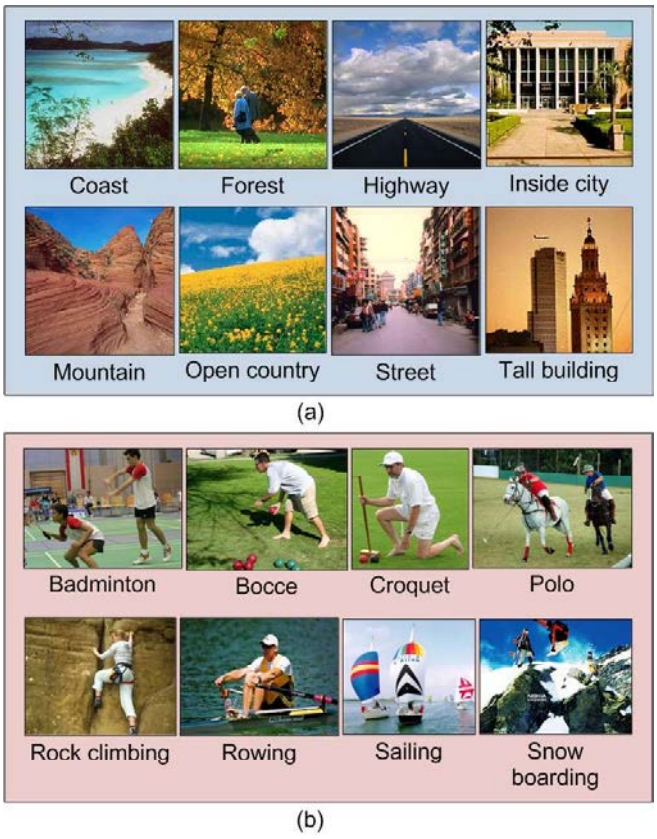

Fig. 2: Some sample images from the (a) MIT Scene dataset, (b) the UIUC Sports Event dataset.

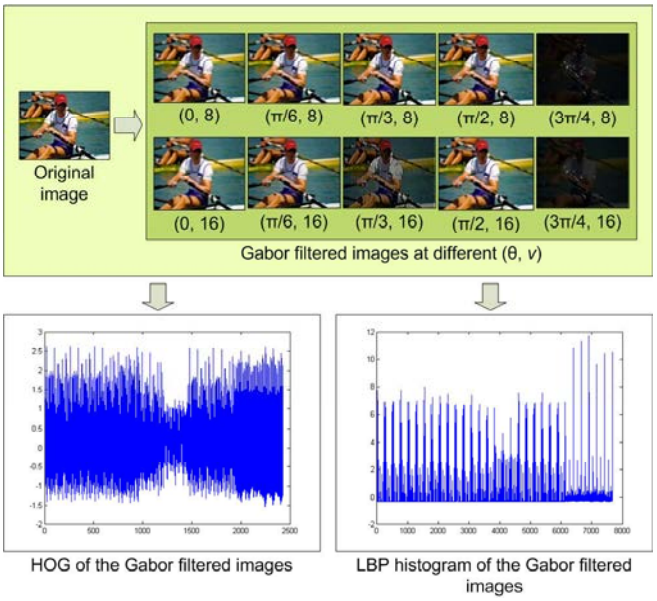

Fig. 3: The generation of the Gabor-HOG and the Gabor-LBP descriptors.

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]. 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 [12]. The EFM feature extraction method has achieved good success for the task of image representation and retrieval [11].

3. Gabor-Based Novel Color Descriptors for Image Classification

This section briefly reviews the color spaces in which our new descriptors are defined, and then discusses the proposed novel descriptors and classification methodology for object and scene 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 [1]. 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°) [3], and is defined as follows:

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

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

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

where the R,G,B values are scaled to $[0,1]$. The oRGB color space [24] 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 $[0,1]$.

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

The Discriminating Color Space (DCS), is derived from the RGB color space by means of discriminant analysis [25]. 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 $\mathscr{X}$ be the 3-D vector in the RGB color space

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

The DCS [15] 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

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

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 [25]. 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 $J_1 = tr(S_w^{-1}S_b)$ [25]. This criterion is maximized when W_D^t consists of the eigenvectors of the matrix $S_w^{-1}S_b$ [25]. Figure 4 shows the grayscale and the color components

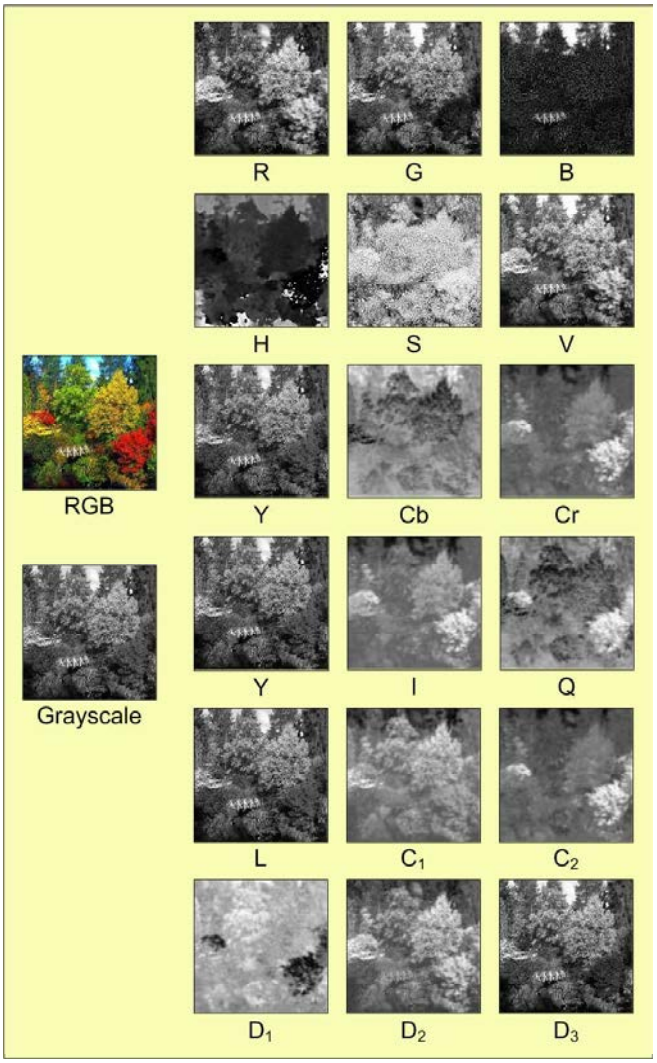

Fig. 4: A sample image from the MIT Scene dataset (labeled RGB) is shown split up into various color components.

of a sample image in the six color spaces used by us in this paper.

3.2 The Gabor-HOG, Gabor-LBP and GLH Descriptors

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

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

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 [21]. Considering that the Gabor wavelet representation captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity [27], we used multi-resolution and multi-orientation Gabor filtering for subsequent extraction of feature vectors. We subject each of the three color components of the image to ten combinations of Gabor filters with two scales (spatial frequencies) and five orientations. For our experiments, we choose $\phi = 0$, $\sigma = 2$, $\gamma = 0.5$, $\theta = [0, \pi/6, \pi/3, \pi/2, 3\pi/4]$, and $\nu = [8, 16]$. LBP histograms of each component of the Gabor filtered images are computed and concatenated to form the Gabor-LBP descriptor. We derive the HOG of the components of the resultant filtered images to produce a novel Gabor-HOG image descriptor. Figure 3 illustrates the creation of the Gabor-HOG and Gabor-LBP features. Each of the two components undergo a dimensionality reduction phase through 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 then integrated to develop the new GLH feature vector, the performance of which is measured on six different color spaces, namely RGB, HSV, oRGB, YCbCr, YIQ and DCS as well as on grayscale. The six GLH color descriptors are further combined to form the Fused Color GLH (FC-GLH) descriptor. Figure 6 gives a comparison of the two descriptors and their fusion for image classification in the three datasets used for our experiments.

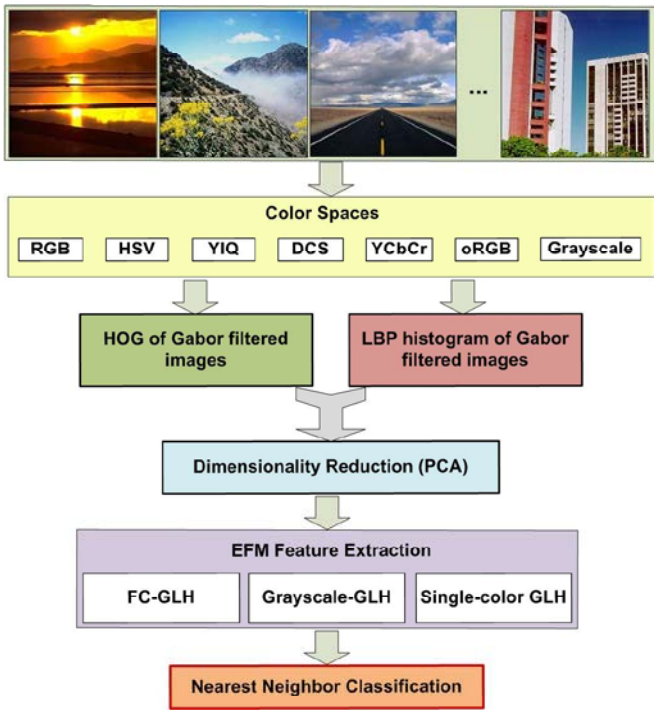

Fig. 5: An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages.

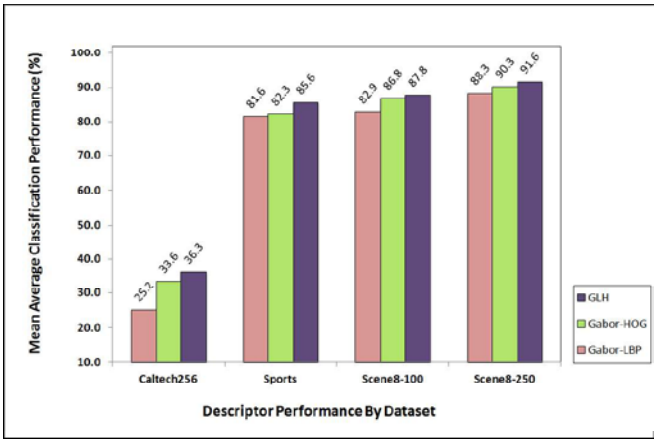

Fig. 6: The comparative mean average classification performance of the FC-Gabor-LBP, FC-Gabor-HOG and FC-GLH descriptors on the Caltech 256, UIUC Sports Event and MIT Scene (with 100 and 250 training images per class) datasets.

Table 1: Comparison of the Classification Performance (%) with Other Methods on the MIT Scene Dataset

#train	#test	GLH		[2]		[26]
2000	688	RGB	**90.3**	CLF	86.4	-
		YIQ	**90.8**	CGLF	86.6	
		FC	**91.6**	CGLF+PHOG	89.5	
800	1888	YIQ	**86.5**	CLF	79.3	
		RGB	**87.3**	CGLF	80.0	
		FC	**87.8**	CGLF+PHOG	84.3	83.7

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

#train	#test	GLH		[4]	
15360	5120	FC	**37.7**	-	
12800	6400	YCbCr	**33.1**	oRGB-SIFT	23.9
		YIQ	**33.7**	CSF	30.1
		FC	**36.3**	CGSF	35.6

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 GLH 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 5 gives an overview of multiple feature fusion methodology, the EFM feature extraction method, and the classification stages.

4. Experimental results

4.1 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 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. The average size of each image is 351x351 pixels. Figure 1 shows some sample images from this dataset.

Table 3: Comparison of the Classification Performance (%) with Other Methods on the UIUC Sports Event Dataset

#train	#test	GLH		[28]		[29]	
560	480	DCS	81.5	HMP	**85.7**	OB	76.3
		RGB	82.9	SIFT+SC	82.7		
		FC	85.6				

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]. In this dataset, YIQ-GLH performs the best among single-color descriptors giving 33.7% success followed by YCbCr-GLH and oRGB-GLH with 33.1% and 32.9% classification rates respectively. Figure 7 shows the success rates of the GLH descriptors for this dataset. The FC-GLH descriptor here achieves a success rate of 36.3%. Table 2 compares our results with those of SIFT-based methods.

4.2 MIT Scene Dataset

The MIT Scene dataset [26] 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. See figure 2(a). 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.

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 also YIQ-GLH is the best single-color descriptor at 90.8% followed closely by RGB-GLH and HSV-GLH. The combined descriptor FC-GLH gives a mean average performance of 91.6%. See Figure 8 for details. Table 1 compares our result with that of other methods. Table 4 shows the class wise classification rates for this dataset on applying the proposed GLH descriptors.

4.3 UIUC Sports Event Dataset

The UIUC Sports Event dataset [31] 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),

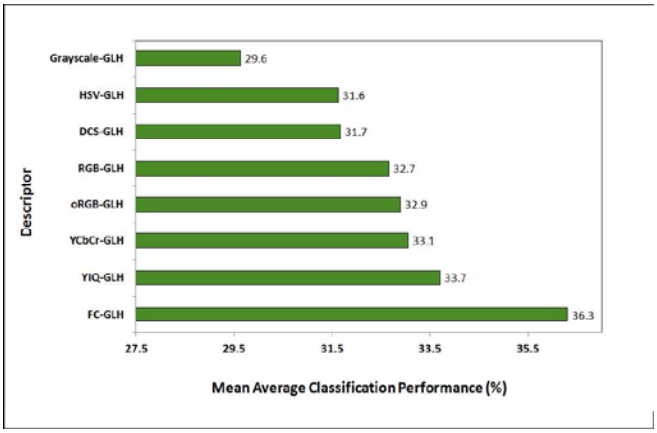

Fig. 7: The mean average classification performance of the proposed GLH descriptor in individual color spaces as well as after fusing them on the Caltech 256 dataset.

Table 4: Category wise descriptor performance (%) on the MIT Scene dataset. Note that the categories are sorted on the FC-GLH results

Category	FC	YIQ	RGB	HSV	DCS	oRGB	YCbCr	Grayscale
forest	**98**	97	97	97	**98**	96	**98**	**98**
street	93	94	93	90	**95**	91	92	91
inside city	**93**	**93**	92	92	91	92	91	89
tall building	**93**	90	90	91	90	92	91	90
coast	**93**	**93**	92	92	89	91	92	91
highway	**92**	**92**	90	**92**	90	90	86	90
mountain	**90**	88	88	87	88	88	86	84
open country	**81**	79	80	79	79	79	80	76
Mean	**91.6**	**90.8**	**90.3**	**90.1**	**90.1**	**90.0**	**89.6**	**88.7**

and rock climbing (194 images). A few sample images of this dataset can be seen in figure 2 (b).

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 [28], [29] have also reported using the same number of images for training and testing. Here RGB-GLH is the best single-color descriptor at 82.9% followed by DCS-GLH, Grayscale-GLH and HSV-GLH respectively. The combined descriptor FC-GLH gives a mean average performance of 85.6%. See Figure 9 for details. Table 3 compares our result with that obtained by other researchers. The FC-GLH descriptor contains RGB, HSV, YIQ, YCbCr, oRGB and DCS-GLH descriptors. The category wise recognition performance of our GLH descriptors on this dataset is shown in table 5.

5. Conclusion

We have presented a new Gabor-based feature extraction method inspired by HOG for color images and combined it with Gabor-LBP features to propose the new GLH descriptor for scene and object image classification. Experimental results carried out using three grand challenge datasets show that our GLH descriptor improves recognition performance over Gabor-LBP and Gabor-HOG descriptors. The color GLH descriptors beat the classification performance of the Grayscale-GLH descriptor in most cases which show information contained in color images can be significantly more useful than that in grayscale images for classification. The fusion of multiple color GLH descriptors (FC-GLH) achieves significant increase in the classification performance over individual color GLH descriptors, which indicates that various color GLH descriptors are not redundant for image classification tasks.

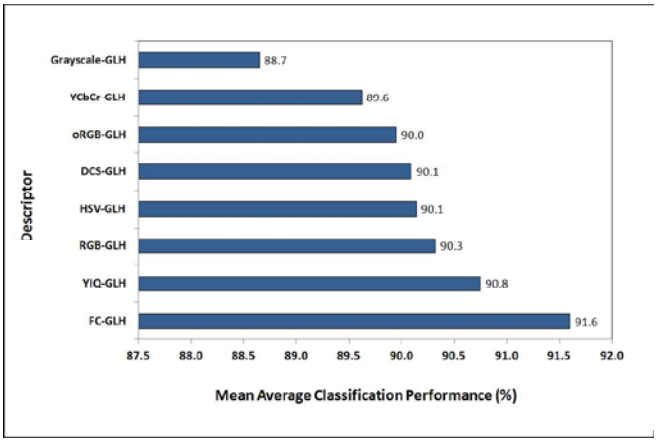

Fig. 8: The mean average classification performance of the proposed GLH descriptor in individual color spaces as well as after fusing them on the MIT Scene dataset.

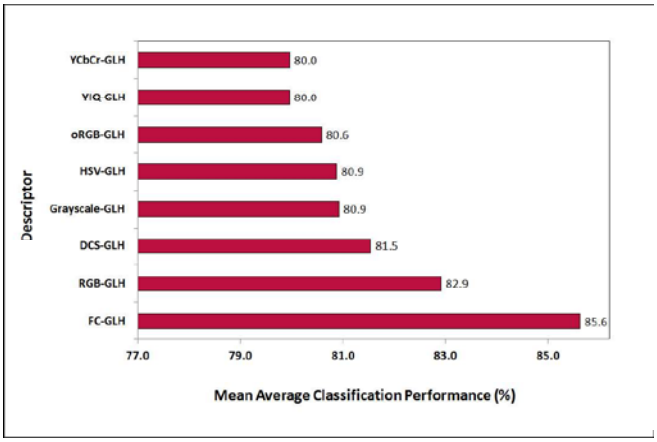

Fig. 9: The mean average classification performance of the proposed GLH descriptor in individual color spaces as well as after fusing them on the UIUC sports event dataset.

Table 5: Category wise descriptor performance (%) on the UIUC Sports Event dataset. Note that the categories are sorted on the FC-GLH results

Category	FC	RGB	DCS	Grayscale	HSV	oRGB	YIQ	YCbCr
rock climbing	95	95	95	94	93	93	**96**	93
badminton	**93**	91	89	91	89	89	88	91
sailing	92	**93**	92	92	90	88	90	91
rowing	**91**	87	89	87	87	89	88	86
snowboarding	**90**	82	84	80	84	84	78	79
polo	87	**89**	78	82	86	74	73	78
croquet	**83**	78	78	71	74	78	78	75
bocce	**55**	48	48	50	44	50	48	47
Mean	**85.6**	**82.9**	**81.5**	**80.9**	**80.9**	**80.6**	**80.0**	**80.0**

References

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

[2] 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.

[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] 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.

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

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

[12] 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.

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

[15] 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.

[16] 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.

[17] 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.

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

[19] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, W. R.P., and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computers*, vol. 42, pp. 300–311, 1993.

[20] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

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

[22] 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 (ICCV'05) Volume 1 - Volume 01*, Washington, DC, USA, 2005, pp. 786–791.

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

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] E. P. X. Li-Jia Li, Hao Su and L. Fei-Fei, "Object bank: A high-level image representation for scene classification & semantic feature sparsification," in *Neural Information Processing Systems (NIPS)*, Vancouver, Canada, December 2010.

[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] L. Fei-Fei and L.-J. Li, "What, Where and Who? Telling the Story of an Image by Activity Classification, Scene Recognition and Object Categorization," *Studies in Computational Intelligence- Computer Vision*, pp. 157–171, 2010.