Novel Gabor-PHOG Features for Object and Scene Image Classification

Atreyee Sinha*, Sugata Banerji, and Chengjun Liu

Department of Computer Science,
New Jersey Institute of Technology,
Newark, NJ 07102, USA
{as739,sb256,chengjun.liu}@njit.edu
http://cs.njit.edu/liu

Abstract. A new Gabor-PHOG (GPHOG) descriptor is first introduced in this paper for image feature extraction by concatenating the Pyramid of Histograms of Oriented Gradients (PHOG) of all the local Gabor filtered images. Next, a comparative assessment of the classification performance of the GPHOG descriptor is made in six different color spaces, namely the RGB, HSV, YCbCr, oRGB, DCS and YIQ color spaces, to propose the novel YIQ-GPHOG and the YCbCr-GPHOG feature vectors that perform well on different object and scene image categories. Third, a novel Fused Color GPHOG (FC-GPHOG) feature is presented by integrating the PCA features of the six color GPHOG descriptors for object and scene image classification, with applications to image search and retrieval. Finally, the Enhanced Fisher Model (EFM) is applied for discriminatory feature extraction and the nearest neighbor classification rule is used for image classification. The effectiveness of the proposed feature vectors for image classification is evaluated using two grand challenge datasets, namely the Caltech 256 dataset and the MIT Scene dataset.

Keywords: Gabor-PHOG (GPHOG), YIQ-GPHOG, YCbCr-GPHOG, FC-GPHOG, PCA, EFM, color spaces, image search.

1 Introduction

Color images provide powerful discriminating information than grayscale images [1], and color based image search can be very effective for face, object, scene, and texture image classification [2], [3], [4]. 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 histogram and local invariant features provide varying degrees of success against image variations such as rotation, viewpoint and lighting changes, clutter and occlusions [5]. Shape and local features also provide important cues for content based image classification and retrieval. Local object shape and the spatial layout of the shape within an image can be described

* Corresponding author.

G.L. Gimel' farb et al. (Eds.): SSPR & SPR 2012, LNCS 7626, pp. 584–592, 2012.

by the Pyramid of Histograms of Oriented Gradients (PHOG) descriptor [6]. Several researchers have described the biological relevance and computational properties of Gabor wavelets for image analysis [7], [8]. Lately, Donato et al. [9] showed experimentally that the Gabor wavelet representation is optimal for classifying facial actions.

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 [7]. In this paper, we design several novel feature vectors based on Gabor filters. Specifically, we first introduce 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 combination of Gabor filters in different orientations. We then measure the classification performance of our GPHOG descriptor on six different color spaces and propose the novel YIQ-GPHOG and the YCbCr-GPHOG features. We further extend this concept by integrating the Principal Component Analysis (PCA) features of the six color GPHOG vectors to produce the novel Fused Color GPHOG (FC-GPHOG) descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) [10], and image classification is based on the nearest neighbor classification rule. Finally, the effectiveness of the proposed descriptors for image classification is evaluated using two datasets: the Caltech 256 grand challenge dataset and the MIT Scene dataset.

2 Novel Gabor-PHOG Features for Object and Scene 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 image classification.

2.1 Color Spaces

A color image contains three component images. The commonly used color space is the RGB color space, from which other color spaces are derived by means of either linear or nonlinear transformations. The HSV color space is motivated by human vision system as 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]. 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. The oRGB color space [11] 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

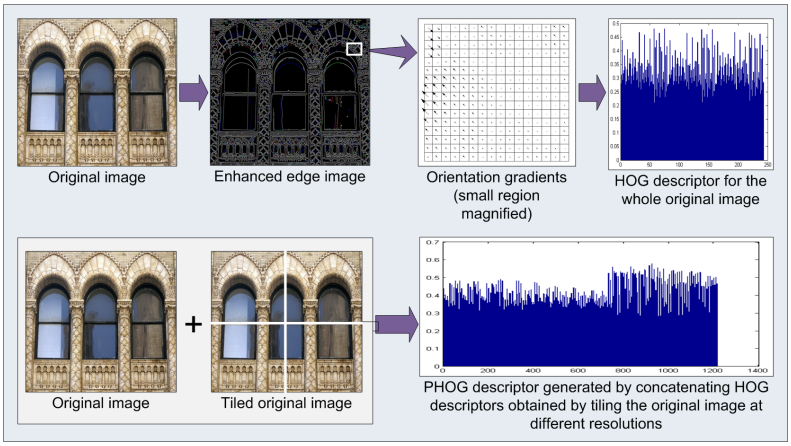

Fig. 1. Generation of the PHOG descriptor for a color image

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 $[0, 1]$. The Discriminating Color Space (DCS) [12], 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. The DCS defines discriminating component images via a linear transformation $W_D \in \mathbb{R}^{3\times3}$ from the RGB color space. The transformation matrix $W_D \in \mathbb{R}^{3\times3}$ may be derived through a procedure of discriminant analysis [13] and has been discussed in [12].

2.2 The Color Gabor-PHOG (GPHOG) and FC-GPHOG Image 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:

$$g_{\nu,\theta,\phi,\sigma,\gamma}(x, y) = \exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}) \exp(i(2\pi\nu x' + \phi)) \tag{1}$$

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 [14].

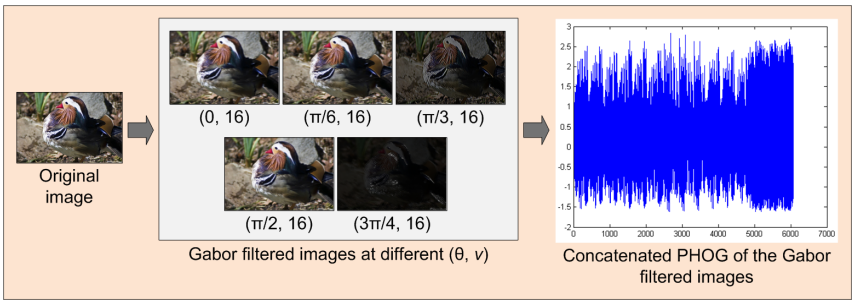

Fig. 2. The generation of the proposed Gabor-PHOG descriptor

The Pyramid of Histograms of Oriented Gradients (PHOG) [6] descriptor, inspired from the Histograms of Oriented Gradients (HOG) [15] and the image pyramid representation of Lazebnik et al. [16], 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]. Figure 1 illustrates the generation of the PHOG feature vector.

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 combinations of Gabor filters. For our experiments, we choose the parameter values as $\phi = 0$, $\sigma = 2$, $\gamma = 0.5$, and $\theta = [0, \pi/6, \pi/3, \pi/2, 3\pi/4]$. We derive the novel Gabor-PHOG (GPHOG) feature vector by concatenating the PHOG of the components of the Gabor filtered images and normalize it to zero mean and unit standard deviation. It should be noted that we computed the PHOG with two levels, and used

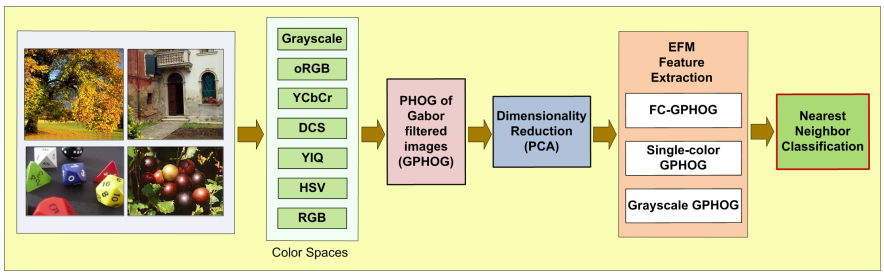

Fig. 3. An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages

Table 1. Comparison of the classification performance (%) with other methods on Caltech 256 dataset. Note that [17] used 250 of the 256 classes with 30 training samples per class.

#train	#test	GPHOG		[4]		[17]
		YCbCr	**30.4**	oRGB-SIFT	23.9	
12800	6400	YIQ	**30.5**	CSF	30.1	
		FC	33.2	CGSF	**35.6**	SPM-MSVM 34.1

$\nu = 16$ as the spatial frequency of the Gabor filters for generating our GPHOG descriptor. Figure 2 illustrates the creation of the proposed GPHOG descriptor.We assess the performance of the GPHOG descriptor on six different color spaces, namely RGB, HSV, oRGB, YCbCr, YIQ and DCS as well as on grayscale and propose two new color feature vctors - the YIQ-GPHOG and the YCbCr-GPHOG descriptors. For fusion, we first use PCA for the optimal representation of our color GPHOG vectors with respect to minimum mean square error. We then combine the PCA features of the six normalized color GPHOG descriptors to form the novel Fused Color GPHOG (FC-GPHOG) descriptor which outperforms the classification results of the individual color GPHOG features.

2.3 The EFM-NN Classifier

Learning and classification are performed using the Enhanced Fisher Linear Discriminant Model (EFM) [10] and the nearest neighbor classification rule. The EFM method first applies Principal Component Analysis (PCA) to reduce the dimensionality of the input pattern vector. The Fisher Linear Discriminant (FLD) is a popular classification method that achieves high separability among the different pattern classes. However, the FLD method, if implemented in an inappropriate PCA space, may lead to overfitting. The EFM method hence 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 GPHOG descriptor and further extracts the EFM features for pattern classification. Similarity score between a training feature vector and a test feature vector is computed using the cosine similarity measure and classification is implemented using the nearest neighbor rule. Figure 3 gives an overview of multiple feature fusion methodology, the EFM feature extraction method, and the classification stages.

3 Experimental Results

3.1 Caltech 256 Dataset

The Caltech 256 dataset [17] holds 30,607 images divided into 256 object categories and a clutter class. The images have high intra-class variability and high

Fig. 4. Some sample images from the Caltech 256 dataset

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 4 shows some sample images from this dataset.

For each class, we choose 50 images for training and 25 images for testing. The data splits are the ones provided on the Caltech website [17]. In this dataset, YIQ-GPHOG performs the best among single-color descriptors giving 30.5% success followed by YCbCr-GPHOG with 30.4% classification rate. Figure 5 shows the success rates of the GPHOG descriptors for this dataset. The FC-GPHOG descriptor here achieves a success rate of 33.2%. Table 1 compares our results with other methods.

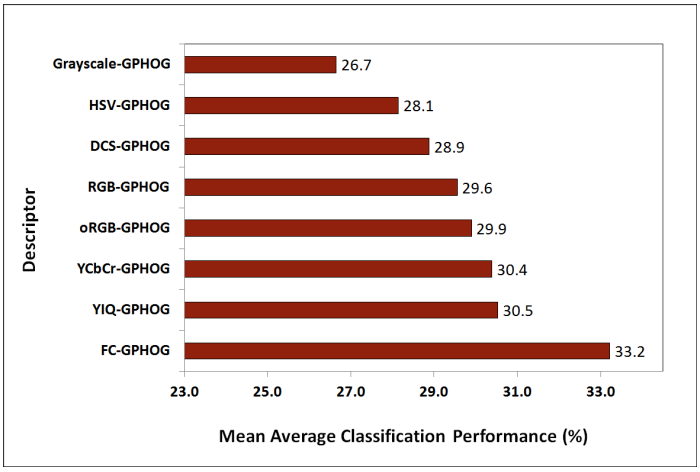

Fig. 5. The mean average classification performance of the proposed color GPHOG and FC-GPHOG descriptors on the Caltech 256 dataset

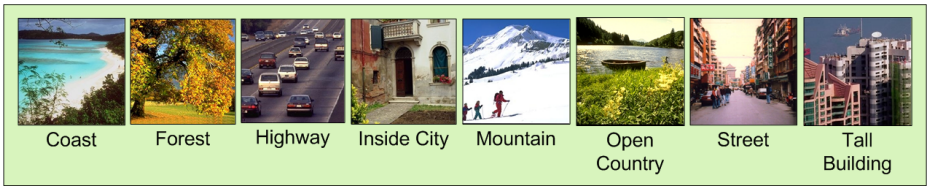

Fig. 6. Some sample images from the MIT Scene dataset

Table 2. Comparison of the classification performance (%) with other methods on the MIT Scene dataset

#train	#test	GPHOG		[2]		[18]
		YCbCr	**87.9**	CLF	86.4	-
2000	688	YIQ	**88.0**	CGLF	86.6	
		FC	**90.2**	CGLF+PHOG	89.5	
		YIQ	**84.6**	CLF	79.3	
800	1888	YCbCr	**84.7**	CGLF	80.0	
		FC	**86.6**	CGLF+PHOG	84.3	83.7

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

Category	FC	YIQ	YCbCr	RGB	DCS	oRGB	HSV	Grayscale
forest	97	96	96	96	96	96	**98**	97
coast	**94**	91	89	91	89	90	88	87
mountain	**91**	88	85	89	87	88	88	83
inside city	**91**	88	90	89	88	86	86	87
highway	90	90	90	88	**92**	90	88	86
street	**90**	88	**90**	89	89	86	85	84
tall building	**90**	88	87	88	86	88	88	86
open country	**79**	75	75	73	75	74	72	68
Mean	**90.2**	**88.0**	**87.9**	**87.8**	**87.7**	**87.3**	**86.5**	**84.8**

3.2 MIT Scene Dataset

The MIT Scene dataset [18] 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 6. 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 too, YIQ-GPHOG is the best single-color descriptor at 88.0% followed closely by

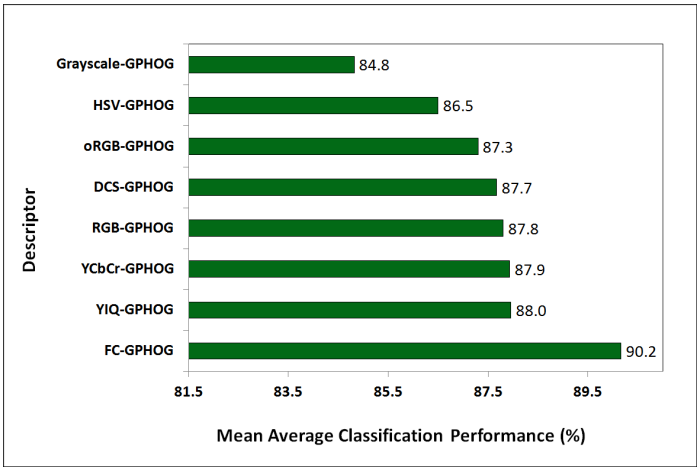

Fig. 7. The mean average classification performance of the proposed GPHOG descriptor in individual color spaces as well as after fusing them on the MIT Scene dataset

YCbCr-GPHOG at 87.9%. The combined descriptor FC-GPHOG gives a mean average performance of 90.2%. See Figure 7 for details. Table 2 compares our result with that of other methods. Table 3 shows the class wise classification rates for this dataset on applying the proposed GPHOG descriptors.

4 Conclusion

The contributions of this paper are in the generation of several novel descriptors for object and scene image classification based on Gabor wavelet transformation. We have introduced a new Gabor-PHOG descriptor and further proposed the robust YIQ-GPHOG and YCbCr-GPHOG features. The six color GPHOG features beat the recognition performance of the Grayscale-GPHOG descriptor which show information contained in color images can be significantly more useful than that in grayscale images for classification. Experimental results using two datasets, the Caltech 256 object categories dataset and the MIT Scene dataset, show that the proposed novel FC-GPHOG image descriptor exceeds or achieves comparable performance to some of the best performance reported in the literature for object and scene image classification.

References

1. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice-Hall (2001)
2. Banerji, S., Verma, A., Liu, C.: Novel color LBP descriptors for scene and image texture classification. In: 15th International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada, July 18-21 (2011)

3. Shih, P., Liu, C.: Comparative assessment of content-based face image retrieval in different color spaces. International Journal of Pattern Recognition and Artificial Intelligence 19(7) (2005)
4. Verma, A., Banerji, S., Liu, C.: A new color SIFT descriptor and methods for image category classification. In: International Congress on Computer Applications and Computational Science, Singapore, December 4-6, pp. 819–822 (2010)
5. Burghouts, G., Geusebroek, J.M.: Performance evaluation of local color invariants. Computer Vision and Image Understanding 113, 48–62 (2009)
6. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: International Conference on Image and Video Retrieval, Amsterdam, The Netherlands, July 9-11, pp. 401–408 (2007)
7. Marcelja, S.: Mathematical description of the responses of simple cortical cells. Journal of the Optical Society of America 70, 1297–1300 (1980)
8. Daugman, J.: Two-dimensional spectral analysis of cortical receptive field profiles. Vision Research 20, 847–856 (1980)
9. Donato, G., Bartlett, M., Hager, J., Ekman, P., Sejnowski, T.: Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(10), 974–989 (1999)
10. Liu, C., Wechsler, H.: Robust coding schemes for indexing and retrieval from large face databases. IEEE Transactions on Image Processing 9(1), 132–137 (2000)
11. Bratkova, M., Boulos, S., Shirley, P.: oRGB: A practical opponent color space for computer graphics. IEEE Computer Graphics and Applications 29(1), 42–55 (2009)
12. Liu, C.: Learning the uncorrelated, independent, and discriminating color spaces for face recognition. IEEE Transactions on Information Forensics and Security 3(2), 213–222 (2008)
13. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press (1990)
14. Lee, H., Chung, Y., Kim, J., Park, D.: Face Image Retrieval Using Sparse Representation Classifier with Gabor-LBP Histogram. In: Chung, Y., Yung, M. (eds.) WISA 2010. LNCS, vol. 6513, pp. 273–280. Springer, Heidelberg (2011)
15. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), Washington, DC, USA, vol. 1, pp. 886–893 (2005)
16. Lazebnik, S., Schmid, C., Ponce, J.: 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, Washington, DC, USA, vol. 2 (2006)
17. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology (2007)
18. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42(3), 145–175 (2001)