Gabor-Based Novel Local, Shape and Color Features for 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. This paper introduces several novel Gabor-based local, shape and color features for image classification. First, a new Gabor-HOG (GHOG) descriptor is proposed for image feature extraction by concatenating the Histograms of Oriented Gradients (HOG) of all the local Gabor filtered images. The GHOG descriptor is then further assessed in six different color spaces to measure classification performance. Finally, a novel Fused Color GHOG (FC-GHOG) feature is presented by integrating the PCA features of the six color GHOG descriptors that performs well on different object and scene image categories. 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 GHOG and FC-GHOG feature vectors is evaluated using two grand challenge datasets, namely the Caltech 256 dataset and the MIT Scene dataset.

Keywords: The Gabor-HOG (GHOG) descriptor, Fused Color GHOG (FC-GHOG) descriptor, Histograms of Oriented Gradients (HOG), Gabor filters, Principal Component Analysis (PCA), Enhanced Fisher Model (EFM), Color spaces, Image search.

1 Introduction

Color contains more discriminating information than grayscale images [1], and color based image search can be very effective for image classification tasks [2], [3], [4]. Some desirable properties of 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], [6]. Shape and local features also provide important cues for content based image classification and retrieval. Local object appearance and shape within an image can be

* Corresponding author.

T. Huang et al. (Eds.): ICONIP 2012, Part III, LNCS 7665, pp. 299–306, 2012.

described by the Histograms of Oriented Gradients (HOG) that stores distribution of edge orientations within an image [7]. Several researchers have described the biological relevance and computational properties of Gabor wavelets for image analysis [8], [9]. Lades et al. [10] used Gabor wavelets for face recognition using the Dynamic Link Architecture (DLA) framework. Lately, Donato et al. [11] showed experimentally that the Gabor wavelet representation is optimal for classifying facial actions.

The motivation behind this work lies in the concept of how people understand and recognize 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 [8]. 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 (GHOG) descriptor by concatenating the Histograms of Oriented Gradients (HOG) of the components of the images produced by the result of applying Gabor filters in different scales and orientations. We then assess our GHOG feature vector in six different color spaces and propose several new color GHOG feature representations. We further extend this concept by integrating the six color GHOG features using a fusion technique that implements feature extraction by means of PCA to produce the novel Fused Color GHOG (FC-GHOG) descriptor. Discriminatory feature extraction applies the Enhanced Fisher Model (EFM) [12], and image classification is based on the nearest neighbor classification rule. Finally, the effectiveness of the proposed descriptors and classification method is evaluated using two datasets: the Caltech 256 grand challenge image dataset and the MIT Scene dataset.

2 Gabor-Based Novel Local, Shape and Color Features for Image Classification

This section discusses the proposed novel descriptors and classification methodology for image classification.

2.1 The Gabor-HOG (GHOG) and Fused Color GHOG (FC-GHOG) 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:

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

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.

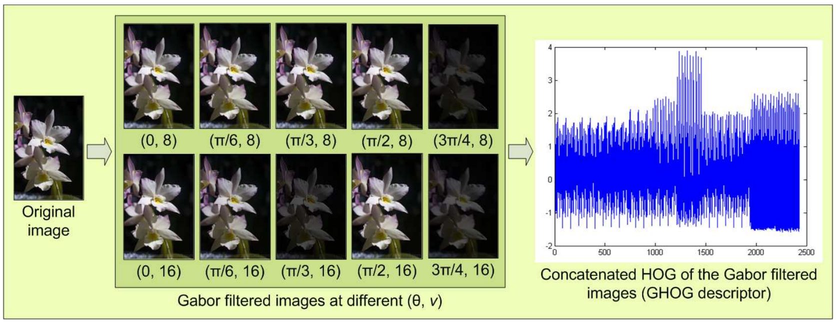

Fig. 1. The generation of the proposed GHOG descriptor

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. Considering that the Gabor wavelet representation captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity [13], [14] 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 $\nu = [8, 16]$.

We concatenate the HOG of the color components of the resultant filtered images and normalize to zero mean and unit standard deviation to produce a new Gabor-HOG (GHOG) image descriptor. Figure 1 illustrates the creation of the GHOG feature, the performance of which is measured on six different color spaces, namely RGB, HSV, oRGB [15], YCbCr, YIQ and DCS [16] as well as on grayscale. Figure 2 shows the grayscale and the color components of a sample

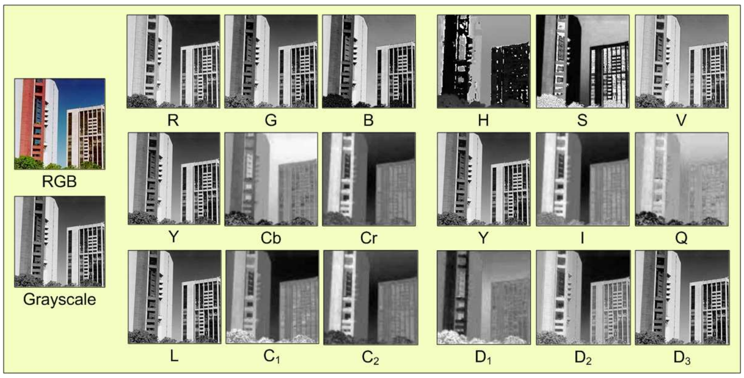

Fig. 2. A sample image from the MIT Scene dataset (labeled RGB) is shown split up into various color components of the RGB, HSV, YCbCr, YIQ, oRGB and DCS

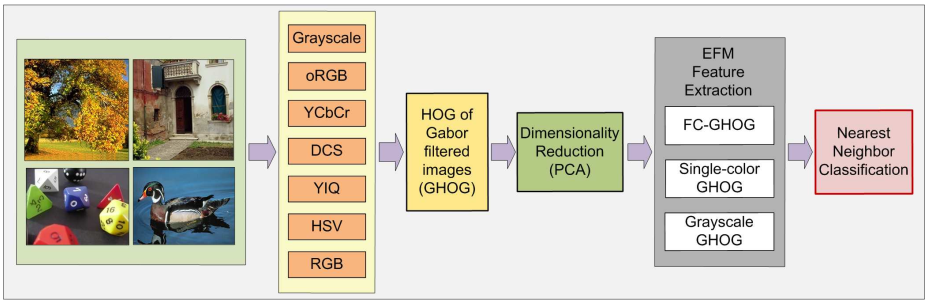

Fig. 3. An overview of multiple features fusion methodology, the EFM feature extraction method, and the classification stages

image in the six color spaces used by us in this paper. For fusion, we first use PCA for the optimal representation of our color GHOG vectors with respect to minimum mean square error. We then integrate the PCA features of the six normalized color GHOG descriptors to form the novel Fused Color GHOG (FC-GHOG) descriptor which outperforms the classification results of the individual color GHOG features.

2.2 The EFM-NN Classifier

We perform learning using Enhanced Fisher Linear Discriminant Model (EFM) [12] and classification is implemented using the nearest neighbor 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, which applies an eigenvalue spectrum analysis criterion to choose the number of principal components to avoid overfitting, improves the generalization performance of the FLD. The EFM method thus derives an appropriate low dimensional representation from the GHOG 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 classification

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	GHOG		[4]		[17]	
		YCbCr	**30.2**	oRGB-SIFT	23.9		
12800	6400	YIQ	**30.7**	CSF	30.1		
		FC	33.6	CGSF	**35.6**	SPM-MSVM	34.1

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

is performed 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 object location variability. Each category contains at least 80, 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

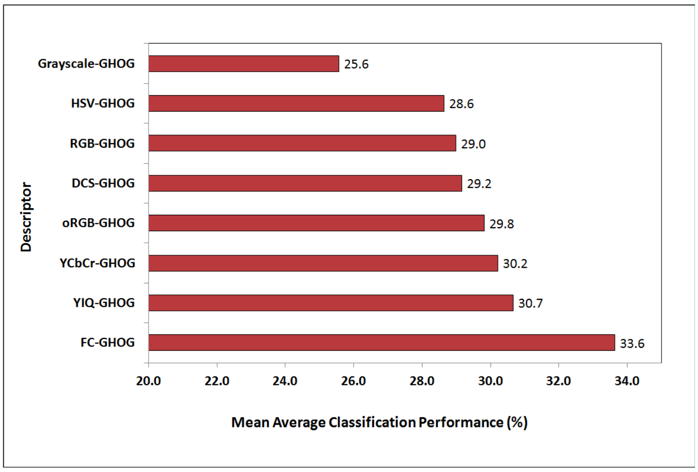

Fig. 5. The mean average classification performance of the proposed GHOG descriptor in individual color spaces as well as after fusing them on the Caltech 256 dataset

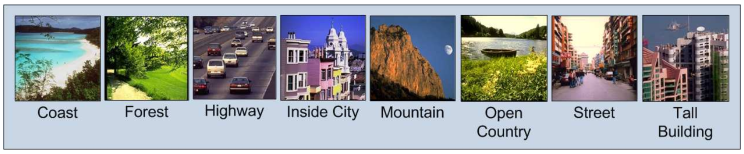

Fig. 6. Some sample images from the MIT Scene dataset

size of each image is 351×351 pixels. Figure 4 shows some sample images from this dataset.

For each class, we use 50 images for training and 25 images for testing. The data splits are the ones that are provided on the Caltech website [17]. In this dataset, YIQ-GHOG performs the best among single-color descriptors giving 30.7% success followed by YCbCr-GHOG with 30.2% classification rate. Figure 5 shows the success rates of the GHOG descriptors for this dataset. The FC-GHOG descriptor here achieves a success rate of 33.6%. Table 1 compares our results with those of SIFT-based methods.

3.2 MIT Scene Dataset

The MIT Scene dataset [18] has 2,688 images classified as eight categories: coast, forest, mountain, open country, highway, inside of cities, tall buildings, and streets. See figure 6 for some sample images from this dataset. All of the images are in color, in JPEG format, and of size 256×256 pixels. There is a large variation in light and angles along with a high intra-class variation.

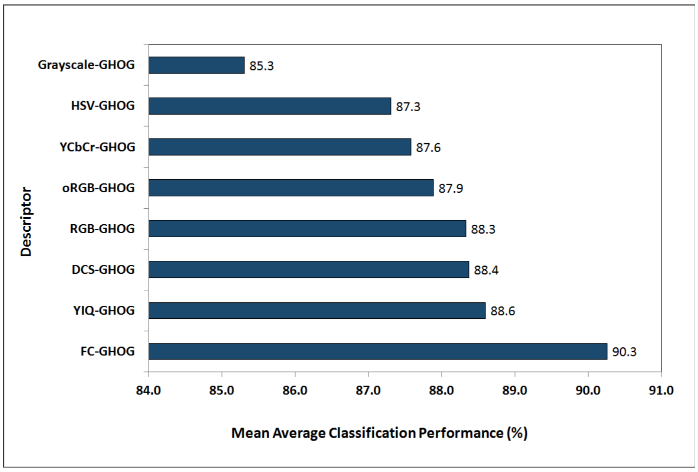

Fig. 7. The mean average classification performance of the proposed GHOG descriptor in individual color spaces as well as after fusing them on the MIT Scene dataset

Table 2. Category wise descriptor (GHOG) performance (%) on the MIT Scene dataset. Note that the categories are sorted on the FC-GHOG results.

Category	**FC**	**YIQ**	**DCS**	**RGB**	**oRGB**	**YCbCr**	**HSV**	**Grayscale**
forest	**98**	**98**	96	97	96	96	97	97
coast	**94**	91	88	90	90	90	88	87
inside city	91	92	**93**	91	**93**	92	90	91
street	90	89	**91**	90	88	88	84	88
tall building	**90**	89	86	87	88	88	87	84
mountain	**90**	86	86	88	87	87	85	79
highway	**88**	86	**88**	**88**	86	82	**88**	84
open country	**81**	77	78	76	77	78	79	73
Mean	**90.3**	**88.6**	**88.4**	**88.3**	**87.9**	**87.6**	**87.3**	**85.3**

Table 3. Comparison of the classification performance (%) with other methods on the MIT Scene dataset

#train	**#test**	**GHOG**		**[2]**		**[18]**
		DCS	**88.4**	CLF	86.4	-
2000	688	YIQ	**88.6**	CGLF	86.6	
		FC	**90.3**	CGLF+PHOG	89.5	
		YIQ	**84.7**	CLF	79.3	
800	1888	RGB	**84.9**	CGLF	80.0	
		FC	**86.9**	CGLF+PHOG	84.3	83.7

From each class, we use 250 images for training and the rest of the images for testing the performance, and we do a five-fold cross validation. Here too, YIQ-GHOG is the best single-color descriptor at 88.6%. DCS-GHOG also performs well to achieve 88.4% success rate. The combined descriptor FC-GHOG gives a mean average performance of 90.3%. See Figure 7 for details. Table 3 compares our result with that of other methods. Table 2 shows the class wise classification rates for the proposed GHOG descriptors on this dataset.

4 Conclusion

We have presented new Gabor-based local, shape and color feature extraction methods inspired by HOG for color images which exceed or achieve comparable performance to some of the best classification performances reported in the literature. Experimental results carried out using two grand challenge datasets show that the fusion of multiple color GHOG descriptors (FC-GHOG) achieves significant increase in the classification performance over individual color GHOG descriptors, which indicates that various color GHOG descriptors are not fully redundant for image classification tasks.

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 (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, 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. Stokman, H., Gevers, T.: Selection and Fusion of Color Models for Image Feature Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), 371–381 (2007)
7. 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)
8. Marcelja, S.: Mathematical Description of the Responses of Simple Cortical Cells. Journal of the Optical Society of America 70, 1297–1300 (1980)
9. Daugman, J.: Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles. Vision Research 20, 847–856 (1980)
10. Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R.P., Konen, W.: Distortion Invariant Object Recognition in the Dynamic Link Architecture. IEEE Transactions on Computers 42, 300–311 (1993)
11. 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)
12. 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)
13. Schiele, B., Crowley, J.: Recognition Without Correspondence Using Multidimensional Receptive Field Histograms. International Journal of Computer Vision 36(1), 31–50 (2000)
14. Liu, C., Wechsler, H.: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
15. 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)
16. 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)
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)