ENHANCED VISUAL CATEGORIZATION PERFORMANCES BY INCORPORATION OF SIMPLE FEATURES INTO BIM FEATURES

Shuangping Huang[1,2], *Lianwen Jin*[1,3]

[1]School of Electronic and Information Engineering,
South China University of Technology
[2]South China Agricultural University,
[3]Key Laboratory of Wireless Communication Networks and Terminals of Guangdong Higher Education Institutes
Guangzhou, 510640, China
E-mail: {huangshuangping, lianwen.jin}@gmail.com

ABSTRACT

Recent studies have demonstrated that Biologically Inspired Model features (BIM) are effective for object or scene categorization. However, according to BIM's forming mechanism, it may not hit the typical pattern due to its blind feature selection. In order to provide more informative pattern information for different visual object classes, a large number of prototypes have to be used in describing images. This leads to huge redundancy which may decrease categorization accuracy. Thus, improving BIM feature's performance by just increasing the number of prototypes is not adequate. In this paper, we propose an integrated approach to address this problem. In our approach, some simple non-biological features such as color histogram and Edge Orientation Histogram (EOH) are incorporated into BIM for discriminative image representation. Experimental results have shown that combination of BIM and simple features can improve visual categorization performances significantly.

Index Terms**—**BIM, simple feature, visual categorization

1. INTRODUCTION

Visual categorization has been a challenging task in computer vision field mainly because of the wide variety of objects to be recognized and the complexity of image backgrounds. In recent years, extensive research and tremendous achievements have been made in this area [1~4]. However, there is still enormous gap between human vision and computer vision on image classifications. Given the vastly superior performance of human vision in this task, it is reasonable to look into biology for inspiration.

As a matter of fact, recent work by Serre et al. [5] has shown that a computational model based on the knowledge of visual cortex can be competitive with the best existing computer vision systems in some of the standard recognition datasets.

However, BIM [5] has its deficiency. It may not hit the typical pattern due to its blind prototype patch selection mechanism. Although a larger number of prototypes can be used to provide more informative pattern information for different visual object classes, many "not so useful" features from background are among those selected. This generates huge redundancy that would not help classification performance.

Some simple features such as color histogram and Edge Orientation Histogram (EOH) [6] can be computed in a fast and simple way. Color is a basic cue for natural scene's classification. Hence it is reasonable to assume that the accuracy of natural scene classification with BIM driven features can be improved by incorporation of color properties. EOH is also important for object and scene recognition because object and scene image often presents strong edges [2]. In fact, EOH is a simple shape descriptor which describes the spatial distribution of edge information. It will refine accuracy of object or man-made scene classification.

In this paper, we propose the incorporation of some additional non-biologically-motivated properties, such as color histogram and EOH into BIM for visual categorization. We have evaluated this approach for both object and scene classification tasks. Experiments are designed on some popular public datasets such as Fei-Fei and Perona [7] dataset (FP), Oliva and Torralba [8] (OT) dataset, Caltech 101[9]. Experimental results show that significant improvements have been achieved in classification performance with our approach.

2. FORMATION OF BIM FEATURES

978-1-4244-7994-8/10/$26.00 ©2010 IEEE

The BIM model consists of four layers of computational units: two S units (S1, S2) and two C units (C1, C2).

Each image containing color information is converted to grayscale. An image pyramid of 10 scales is then created, with each factor being $2^{1/4}$ smaller than the one next to it. The pyramid will enter the subsequent four layers of BIM model and finally C2 descriptors will be formed.

S1 layer is computed by applying a bank of Gabor filters with different orientations but with the same size of 11×11 to the scaled versions of the image. The Gabor filters can be represented by the following equation:

$$G(x,y)=\exp\left(-\frac{\left(X^2+\gamma^2Y^2\right)}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}X\right) \tag{1}$$

where $X = x\cos\theta + y\sin\theta, Y = -x\sin\theta + y\cos\theta$, x and y vary between -5 and 5, θ is $0^0, 45^0, 90^0 or\, 135^0$. Parameters λ (wavelength), σ (effective width) and γ (aspect ratio) are set to be 5.6, 4.5, 0.3 respectively. The response of a patch of pixels X to a particular filter G is given by:

$$R(X,G)=\left|\frac{\sum X_i G_i}{\sqrt{\sum X_i^{\,2}}}\right| \tag{2}$$

Therefore, we obtain a S1 pyramid of 10 layers with each having 4 different orientations.

For each image, C1 units are obtained after a max pooling operation over nearby S1 units of the same orientation and then over larger local regions. Due to the pyramidal structure of S1, we use the same 3D max filter of 10×10 units across in position and 2 units deep in scale. Final C1 units are computed by sub-sampling S1 maps using a cell grid of size 10 with a step of 5 in positions but only 1 in scale, giving a sampling overlap factor of 2 in both position and scale.

Before S2 layer computation, prototype patches representing some typical patterns have to be learned during training. The prototypes are randomly sampled from C1 units of the training images at random position and scale. The patch has four different sizes: 4×4, 8×8, 12×12 and 16×16. Since the prototypes are learned randomly from un-segmented images, many will not actually hit the object of interest, and others may not be useful for the classification task [10].

In S2 layer, afferent C1 units are compared with the stored prototypes as in equation (3). If we have N stored prototypes, S2 pyramid are generated by computing N times across all positions and scales. In essence, the procedure calculates the match of each prototype with C1 units in a traversal way. When there is better match between prototype and C1 units at a certain position and a scale, there is stronger response in S2 layer. The response of a patch of C1 units X to a particular prototype P is given:

$$R(X,P)=\exp\left(\left|\frac{\left\|X-P\right\|^2}{2\sigma^2\alpha}\right|\right) \tag{3}$$

Where standard deviation σ is set to be 1 and normalization factor $\alpha = \left(n/4\right)^2$,where n is 4, 8, 12 or 16 corresponding to different patch sizes.

Final shift- and scale-invariant C2 responses are calculated by taking a global maximum over all scales and positions over the entire S2 lattice. We keep only the value of the best match and discard the rest. The result is a C2 vector of N values. Before entering the follow-up process of image representation and final category analysis, C2 is "sphered" as in [10].

3. INTEGRATION OF SIMPLE COLOR AND SHAPE ATTRIBUTES

When computing C2 feature for an image, color information is removed completely. In order to integrate color properties, HSV-based histogram is used in our approach. The computation operates in the way as shown in Figure 1: the image is divided equally into four small regions, and the histograms for H, S, V values over the whole image and the small regions are computed with certain number of bins. All the histograms are concatenated into one vector. It is then normalized by means of "sphere" before concatenation with C2.

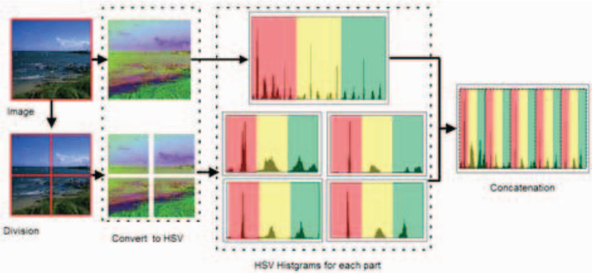

Figure 1. Forming procedure of HSV histogram

Shape information becomes crucial for categorization tasks. This is especially true where there is absence of color. Although there are many complicated shape descriptors which show good invariance, we emphasize a simple and faster, yet robust scheme to help improve BIM performance. Edge Orientation Histogram (EOH) [6] is a concise and quantitative way of describing object shapes. The Canny [11] edge detector is used to retrieve the edge points, and a histogram of the directions of the edge points is used to represent the shape. It is also "sphered" before it is combined with BIM features. Figure 2 shows an example of two different types of scenes, which includes 'Tallbuilding', 'Highway', and their corresponding EOH. Obvious similarity between intra-class images and difference between inter-class images shows EOH's discriminating effect.

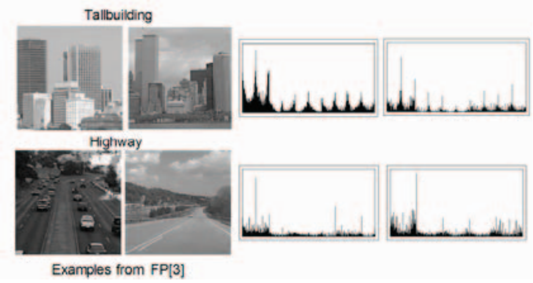

Figure 2. EOH's discriminating effect

4. EXPERIMENTAL EVALUATION

Our ideas are validated by measuring the accuracies of scene and object classification, respectively.

4.1. Tested on scene categorization

FP dataset [7] is available only in grayscale. So it is used to evaluate the effectiveness of EOH incorporation alone. OT [8] is a color scene dataset and is thus used to evaluate the effectiveness of combination of HSV histogram (abbreviated as 'HSVHist') and EOH with BIM features. The classification accuracy reported here is the average precision of all categories which is obtained by taking the average of 8 independent runs. In each run, 100 training images are selected randomly for each class, and the remainder is for testing.

4.1.1. FP dataset

FP dataset contains 13 scene categories. It consists of 3863 images, which include bedroom, kitchen, living-room, office, highway, inside-city, tall-building, suburb, streets, coast, forest, mountain, and open-country.

The 2600 prototypes with which C2 responses are computed are extracted from training images with randomly selected sizes from random locations of C1 response fields. Multi-class SVM with one-versus-all method [12] is used in the evaluation as the classifier.

Table 1 shows the result of classification with "C2" and "C2+EOH". Some benchmark results from recent literatures are also listed in the table for comparison. From the table, we can see that incorporation of EOH leads to approximately 7.5% increase in average precision as compared to that with C2 only. Same is true as we compare our results with that using PLSA method [13], i.e. about 8.2% improvement in average precision. Comparing our results with that from the latest work of Terashima [14], we can see they are very close. It is worth noting that in our method, only one SVM classifying stage is adopted based on simple concatenation of BIM and EOH, while Terashima's work was based on C2 feature but with added C1 histogram and two stages of image analysis. It makes Terashima's classification system much more complicated than ours.

Table 1. Results obtained with "C2", "C2+EOH" and benchmark in recent three years

Methods	Average Accuracy
Only C2	74.79±0.81
C2+ EOH	80.42±0.41
PLSA[13]	74.3±1.3
Method in [14]	81.2±0.4

4.1.2. Color OT dataset

The OT dataset is composed of 2688 color images of 8 categories. Experimental results are obtained in the case of randomly sampling 1600 C2 features.

Figure 3 shows the confusion matrix for the eight categories when "C2", "C2+EOH", "C2+HSVHist" and "C2+EOH+HSVHist" are used. From the figure, it is clear that after incorporation of EOH and HSV histogram into BIM features, the classification accuracy across the board for the scene has been improved significantly, some by up to 28%. Adding EOH or HSV histogram alone results in 6% and 5% improvement in average precisions, respectively.

The average precision for eight categories is listed on the top of confusion matrix corresponding to different methods.

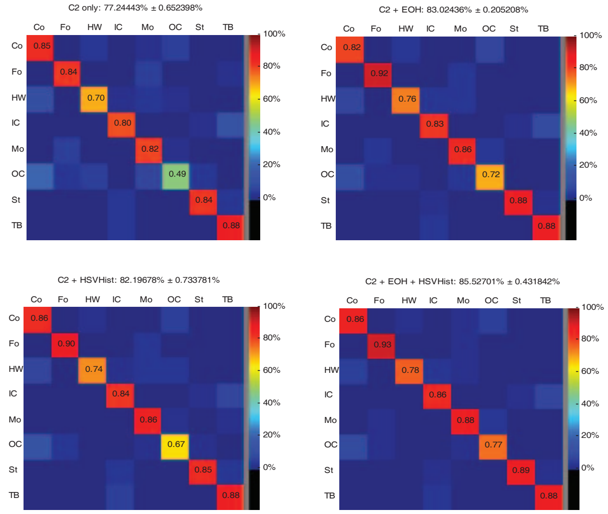

Figure 3. Confusion Matrix on OT
Co: Coast, Fo: Forest, HW: HighWay, IC: InsideCity, Mo: Mountain, OC: OpenCountry, St: Street, TB: TallBuilding

4.2. Tested on object categorization

4.2.1. Caltech 101

CalTech101 contains 9197 images comprising 101 object classes plus a background class. The dataset includes color and grayscale images. So color information is not considered here. To conduct this experiment, we use 3030 features with best parameters learned in two-fold cross validation. In each run, 15 images are sampled randomly for training and the remainder is for testing. For comparison, pairwise SVM [12] with majority voting rule is utilized.

The results are summarized in table 2, along with some literature results. From the table, we can see that integration of EOH in C2 leads to 2.2% improvement in average precision from all of the 101 object categories. In J. Mutch's work [10], even though more prototypes are used, the average precision is lower than what we report here.

Table 2. Results obtained with "C2", "C2+EOH", EBIM, BIM with 4075 features

Methods	Num. of features	Average Accuracy
Only C2	3030	59.94±0.62
C2+ EOH	3030	61.25±0.64
EBIM [15]	1000	49.8 ± 1.25
J. Mutch et al. [10]	4075	51

5. CONCLUSION

In this paper, we proposed a new method of incorporating simple features into BIM driven image representation for categorization performance enhancement. The simple features include HSV histogram and EOH. Experimental results have shown that with our method, higher accuracy can be achieved than that with C2 features only. Meanwhile some classification performance can be competitive with the best reported methods on several public image recognition datasets. Although BIM is a valuable model for categorization, we found that addition of simple histogram features, such as HSV histogram or EOH, can make up for the deficiency of it not hitting typical class pattern. The better classification results obtained by adding color or shape properties suggest that low level image features can improve classification accuracy when they are combined with higher level biological features.

ACKNOWLEDGEMENT

This work is supported in part by the National Science Foundation of China (NSFC) (grants no. U0735004, and no. 60902087), Guangdong National Science Foundation (grant no. 07118074).

REFERENCES

[1] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," In Workshop on Statistical Learning in Computer Vision, ECCV, Prague, Czech Republic, pages 1–22, May 2004.

[2] J. Fan, Y. Gao, H. Luo, and G. Xu, "Statistical modeling and conceptualization of natural images," Pattern Recognition, 38:865–885, 2005.

[3] L. Fei-Fei, R. Fergus, and P. Perona, "A bayesian approach to unsupervised oneshot learning of object categories," ICCV, volume 2, Nice, France, pages 1134–1141, October 2003.

[4] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," ICCV, volume 2, pages 1458–1465, 2005.

[5] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," CVPR, 2005.

[6] David Geronimo, Antonio Lopez, Daniel Ponsa, and Angel D. Sappa, "Haar Wavelets and Edge Orientation Histograms for On-Board Pedestrian Detection," IbPRIA, Part I, LNCS 4477, pp. 418 – 425, 2007.

[7] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," CVPR, VOL. 2, 524–531, 2005.

[8] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," IJCV, 42(3): 145–175, 2001.

[9] F. Li, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," In Workshop on Generative-Model Based Vision, CVPR, page 178, 2004.

[10] Jim Mutch and David G. Lowe, "Multiclass Object Recognition with Sparse, Localized Features," CVPR, pages 11-18, 2006.

[11] J. Canny, "A computational approach to edge detection," PAMI, Volume 8, pp. 679-698, 1986.

[12] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.

[13] Anna Bosch, Andrew Zisserman and Xavier Munoz, "Scene classification using a hybrid generative/discriminative approach," PAMI, 2008.

[14] Yoshito Terashima, "Scene Classification with a Biologically Inspired Method," MIT-CSAIL-TR-2009-020, 2009

[15] Yongzhen Huang, Kaiqi Huang, Liangsheng Wang, Dacheng Tao, Tieniu Tan and Xuelong Li, "Enhanced Biologically Inspired Model," CVPR, 2008.