Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2009-020
CBCL-277

May 10, 2009

Scene Classification with a Biologically Inspired Method

Yoshito Terashima

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Scene Classification with a Biologically Inspired Method

Yoshito Terashima

Abstract

We present a biologically motivated method for scene image classification. The core of the method is to use shape based image property that is provided by a hierarchical feedforward model of the visual cortex [18]. Edge based and color based image properties are additionally used to improve the accuracy. The method consists of two stages of image analysis. In the first stage, each of three paths of classification uses each image property (i.e. shape, edge or color based features) independently. In the second stage, a single classifier assigns the category of an image based on the probability distributions of the first stage classifier outputs. Experiments show that the method boosts the classification accuracy over the shape based model. We demonstrate that this method achieves a high accuracy comparable to other reported methods on publicly available color image dataset.

1. Introduction

The ability to distinguish between multiple semantic categories of natural scene images is a challenging and attractive target for the field of computer vision. Although human vision is still superior to computer vision on image classification tasks, recent studies have made significant advances in improving the ability of computers to classify images. Some basic, although not very accurate, approaches for classifying scene images are on the basis of low level features such as color histogram information and power spectrum [1, 2, 3, 15]. A number of recent studies showed improvement in the accuracy of classification by using previously assigned intermediate representations to classify images [4, 5, 6, 7]. There are several ways to obtain such intermediate representations, for instance, labeling by human observers [4], classifying grids based on color and edge [6], retrieving automatically based on Latent Dirichlet Allocation [5] or probabilistic Latent Semantic Analysis [7] .

A method that does not use intermediate representations, but instead relies on the spatial information of features, has been described [8]. This method applied to 15 scene categories has achieved a significant improvement in the accuracy of classification over a simple bag-of-features approach. The key point of the improvement is to utilize local distribution of edges or SIFT features together with a histogram that shows the number of the features found in a local region. The idea of using local distribution of features has been applied to intermediate representation approaches to achieve state-of-the-art performance [7].

These approaches have achieved high accuracy on scene classification, even though they do not require detection of explicit objects existing in the whole image. Taken together with the well known fact about the human vision that scenes can be viewed in a short time when there is not enough time for eye movements or attention shifts [9], the results of these studies suggest that global properties provide sufficient discrimination power for natural scene images.

If global properties classify scene images well, object recognition mechanisms that incorporate current knowledge about the early stage of human and primate vision analysis [10, 18] may provide good performance on scene classification.

Though it is well known that humans can recognize objects without color, presence of color improves the accuracy and response time in the early stage of scene categorization by human subjects [11]. We therefore assumed that the accuracy of scene classification with the biologically inspired vision model may be improved by including color analysis.

In addition, research on the vision system of primates has shown evidence that the luminance and chrominance components of an image signal are processed through separate pathways [12]. The concept of analyzing color and texture information separately has been applied to improve the accuracy of image classification [3].

In this paper, we describe a method for classifying scene images that separately analyzes three types of image attributes. It builds on the biologically inspired model of shape based analysis and has the potential to incorporate color information and local distribution of edges to refine the accuracy of image classification. The structure of the method is described in section 2 and the results of evaluation are shown in section 3. We demonstrate that this method achieves a high accuracy comparable to other state-of-the-art methods.

2. Approach

Figure 1 shows the outline of the proposed method. The method is composed of two stages of analysis. In the first stage, three paths of classification use different image features independently of each other. The core of the method is shape based analysis which follows the standard model of the visual cortex described in [10]. Edge based analysis and color based analysis are the other two paths in the first stage of the method. In the second stage, outputs from the first stage are combined by different strategies to assign a scene category to the test image. The details of how to compute the shape based feature is described in section 2.1 and how to compute the edge and color based features is described in section 2.2.

2.1. Biologically inspired model

The standard model of vision follows a theory of immediate object recognition (first 100 to 200 milliseconds after the recognition process starts). The model consists of four layers of computational units, namely two S units (S1, S2) and two C units (C1, C2). S and C respectively correspond to 'Simple' and 'Complex' cells discovered by Hubel and Wiesel [13] in the primary visual cortex (V1). The values of C2 units are computed in a hierarchical way thorough S1, C1, S2 and C2 in turn. In the classification stage, the responses of C2 units (C2 feature) of training images are used to learn the category and those of test images are used to predict the category of the image.

The method described in this paper uses the standard model source code that is available on the Web [17]. The parameter values in the model are the same as in [10]; these match what is known about the primate vision system. All input images that have color information are converted to grayscale before the computation of the S1 units.

Responses of S1 units are computed by 2-D Gabor filter using the following equation:

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

where $X = x\cos\theta + y\sin\theta$, $Y = -x\sin\theta + y\cos\theta$ and λ (aspect ratio) = 0.3. The filter sizes are varied from 7×7 to 37×37 pixels in 2 pixel increments. Each size of the filter has a corresponding wavelength λ, and effective width σ (Table 1 in [10] for detail). The orientation θ is either 0°, 45°, 90° or 135°, therefore, S1 units form a total of 64 receptive fields (16 sizes $\times$ 4 orientations).

The computed S1 responses are pooled together in the C1 layers with a MAX operation. This operation brings some tolerance against change in location or size. The C1 layers consist of several scale bands and several orientations of C1 units. Each scale band of C1 units pools together two adjacent sizes of S1 units with the same orientation. In this model, the number of scale bands is 8 and the pooling grid size of scale bands ranges from 8×8 (scale 1) to 22×22 (scale 8) by increments of 2. C1 responses of each scale band are computed with S1 responses subsampled to be fit to the grid size. The pooling grids overlap by the amount of half of a pooling grid size. C1 units form a total of 32 receptive fields (8 sizes $\times$ 4 orientations).

In the S2 layers, S2 units compute the responses between patches of C1 units and previously stored prototypes composed of some C1 units. The prototypes P with the size of $n \times n$ and 4 orientations are extracted randomly from the C1 layers of training images and stored. S2 responses r are represented by the Euclidean distance between the image patches X of C1 units and particular prototypes P. The response of the patch X_j to the prototype P_i is given by the Gaussian radial basis function:

$$r_{ij} = \exp\left(-\left\|X_j - P_i\right\|^2 \big/ 2\sigma^2\right) \quad (2)$$

where σ is a tuning parameter for the sharpness of the response and is the absolute value of a P_i; the square root of its inner product.

C2 units take the maximum over all scales and positions of S2 responses in the previous layer. Therefore, the number of C2 units is the same as the number of prototypes. When the number of prototypes is N, each image has an N-dimensional vector consisting of C2 responses. A C2 feature is represented by the vector and can be considered as a global property because each C2 response does not have scale

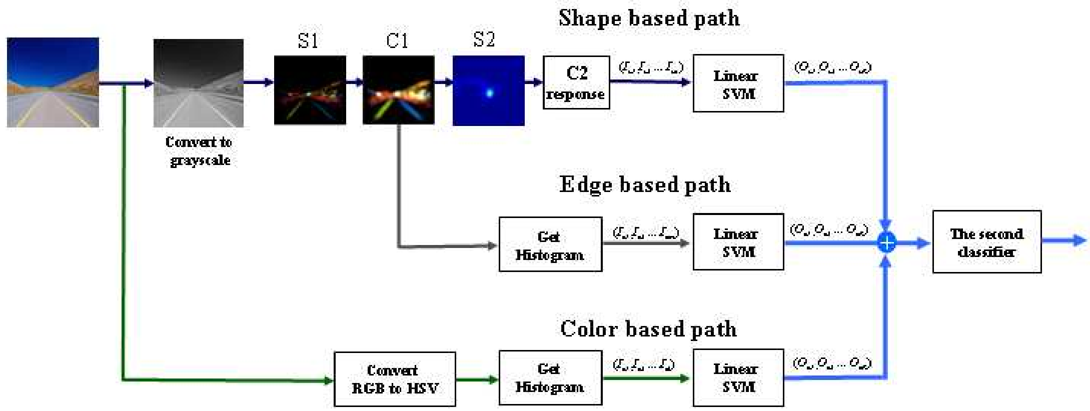


Figure 1. Overview of the proposed method. In the first stage, images are analyzed by three separate paths (only shape based and edge based paths are used in the analysis of grayscale images). In the second stage, the separate first stage outputs are combined by various classifiers to assign the output category for the image.

or position information, but is the value of the strongest S2 response to a particular shape.

The size of a prototype in this paper is chosen among 4, 8, 12 and 16 randomly at a learning stage. We preliminarily examined the possibility that extending the range of prototype sizes may improve the accuracy of scene image classification. The result suggested that adding smaller and larger sizes of prototypes did not seem to improve the accuracy significantly and that the increase in the number of prototypes might even decrease the accuracy.

2.2. Addition of edge and color based features

First, we explain how to compute edge based features. The histograms of C1 responses represent edge information in this system. After an image is divided into several equally sized regions (N_r), C1 histograms are computed for each region so that the histograms have a local distribution of edges. After the responses of S1 and C1 units are computed from each region with the same model parameters as shown in Section 2.1, a histogram with N_b bins is calculated in the each region. As there are N_r regions, there are N_r histograms, and they are concatenated for an image. In this paper, because C1 units have 32 fields (8 sizes $\times$ 4 orientations) in every single region, the dimension of the histogram is $32 \times N_b$ and the total dimension for the image is $32 \times N_b \times N_r$. The values of histograms are normalized by the number of C1 units that are included in the corresponding field where the histogram is calculated so that the range of the histogram is 0 to 1 and different sized images can be analyzed.

Color histogram information for an image has been used for object recognition and image retrieval. Although several types of color descriptors may be used (such as Lab, HSL, LUV, etc), we used HSV information in our method, based on the number of studies which demonstrate its effectiveness in scene image classification [2, 7].

While a color image is converted to grayscale in computing shape based and edge based paths, original data set images are used in the color based path. The image is divided equally into several small regions (N_{rc}), and the histograms for H, S and V values are computed with N_b bins. The value of a histogram is normalized by the number of pixels in the region for the range 0 to 1. Three types of histograms of H, S and V are concatenated into one vector, therefore the dimension of the vector of each region is $3 \times N_{bc}$ and the total of the whole image is $3 \times N_{bc} \times N_{rc}$.

The features obtained for each path at the first stage analysis of our method (shape based path, edge based path and color based path) are combined in the second stage in the following manner: 1) three multi-class Support Vector Machine (SVM) classifiers are trained separately with C2 features, C1 histograms, HSV histograms using training images, 2) as the first stage of analysis, any test image gets three different category predictions as outputs from the individual SVMs for each path, 3) finally, the first stage outputs are used as the inputs for the second stage classifier which provides the final result of classification for the image.

3. Experimental Evaluation

We evaluate our method by analyzing each path separately in the first stage and in combination at the second stage. The *Caltech Natural Scene image (NS)* dataset [5] is used to evaluate the shape based and edge based paths. The *MIT Color NS* dataset [4] is used to evaluate the full feature method.

3.1. Edge based path (Caltech NS dataset*)*

The performance of the method is evaluated by measuring the accuracy of scene classification. If the proposed method, which combines C2 feature and C1 histogram, works effectively, its accuracy will exceed the accuracy of either of the C2 feature or the C1 histogram alone. The Caltech NS dataset is composed of 13 categories of grayscale images: suburb (241 images), bedroom (174), kitchen (151), living-room (289), office (216), coast (360), forest (328), highway (260), inside-city (308), mountain (410), open-country (260), tall-building (356). The image size is around 300×250 pixels, and the images are not resized in the evaluation.

The accuracy of the 13-category classification reported here is the average of 13 precisions of each category which is obtained by taking the average of five independent runs. In each run, 100 training images are selected randomly, and 50 test images are also selected randomly form the rest of images for each category.

The prototypes (see sec. 2.1), with which the C2 responses are computed, are extracted from training images with randomly chosen sizes from random locations of C1 response fields. Two multi-class SVMs are used as the first stage classifier for the C2 features and the C1 histograms respectively. Each SVM is trained with one-versus-all method and its parameters are determined based on twofold cross validation with all training images. The class or category of a particular test image is designated by taking the class having the highest response of SVM outputs.

Figure 2 shows the result of the classification with the C2 features. The graph in the figure expresses the correlation between the accuracy/deviation and the number of prototypes. While the number of prototypes greatly affects the accuracy when the number is small, accuracy does not improve significantly after the number reaches over 2600 (equivalent to 2 prototypes per training image).

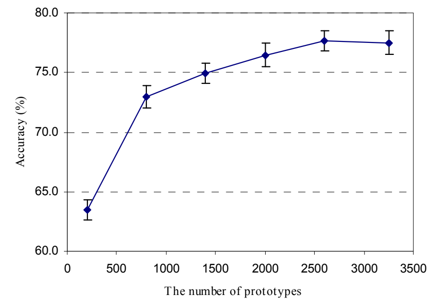


Figure 2. Accuracy and standard deviation versus the number of prototypes on classification with C2 feature.

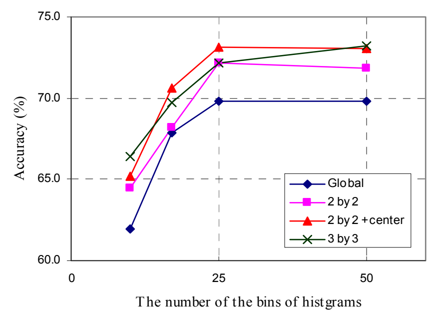


Figure 3. Accuracy versus the number of bins of C1 histogram for several types of region divisions.

Figure 3 shows the result of the classification with the C1 histograms. The graphs express the correlation between the accuracy and the number of bins of C1 histograms over several types of region divisions. In the figure, ‘2 by 2’ has 4 regions in an image, which is divided into 2 regions horizontally and into 2 vertically. Similarly, ‘3 by 3’ has 9 regions, and ‘2 by 2 + center’ has 5 regions with the region which has the same area size as others located at the center in addition to the ‘2 by 2’. In ‘Global’, the entire image is used without any partitioning. This result shows that dividing the image and getting histograms from the small regions improves the accuracy of the classification and that 25 bins is sufficient in most cases. Interestingly, ‘Global’ has the lowest accuracy, suggesting that dividing the image improves the accuracy for the edge based analysis. The best result of 73.2% accuracy is achieved in two conditions: ‘2 by 2 + center’ with 25 bins (5 regions; vector dimension 4000) and ‘3 by 3’ with 50 bins (9 regions; vector dimension 14400). Interestingly, the '2 by 2 + center' provides competitive performance to the '3 by 3' in this case.

Having determined the accuracy of C2 response and of C1 histograms to be 77.7% and 73.2% respectively, the target for the second stage classifier is to get higher accuracy than 77.7%. The inputs for the second stage classifier are the outputs of the first stage classifiers. A softmax function is used to convert them to probabilities [14]. At the second stage of the method, images are classified by combining the first stage outputs using the same strategies as in [16]. 1) Summation: the two 13-dimentional vectors consisting of the first stage classifier outputs are added together to create a new vector. The category that has the maximum output over the 13 elements of the new vector is the predicted category of the image. 2) Product: The procedure is similar to 'Summation', however, instead of taking the sum of each vector element, products for each element are calculated to create a new vector. The category which has the maximum value over the 13 factors is the predicted category. 3) Majority vote: the maximum value of all elements of two 13-dimensional vectors is searched, and the category having the maximum value over 26 elements is the predicted category. Table 1 shows the classification result with our method using the second stage classifiers. 'Summation' and 'Product' strategies successfully improved the accuracy, as compared to the accuracy with only C2 feature.

Table 1. The result of grayscale image classification with C1 histogram and C2 feature.

	5 regions 25 bins	9 regions 50 bins
Summation	80.7 ± 1.1	81.2 ± 0.4
Product	80.6 ± 1.2	81.2 ± 0.5
Majority vote	77.4 ± 1.6	78.9 ± 0.8
C2 feature	77.7 ± 0.9	
C1 histogram	73.2 ± 2.2	73.2 ± 1.1

3.2. Full feature method (MIT NS dataset)

Next, we evaluated the method with full feature including HSV information on the MIT Color NS dataset. The MIT dataset, which has been inherited by the Caltech NS dataset, is composed of 2680 images of 8 categories: coast, forest, highway, inside-city, mountain, open-country and tall-building. Each image is 256×256 pixels.

Table 2. The result of color image classification with full feature method (C1 histogram, HSV histogram and C2 feature)

	Full feature	C2 feature + C1 histogram	C2 feature + HSV histogram
Summation	**88.6** ± 0.2	86.9 ± 0.4	87.0 ± 0.3
Product	85.8 ± 1.0	86.9 ± 0.4	84.7 ± 0.8
Majority vote	88.5 ± 0.2	85.6 ± 0.9	84.7 ± 0.8
Stack	81.9 ± 0.6	84.5 ± 0.4	79.9 ± 1.2
C2 feature	84.1 ± 0.6		
C1 histogram	80.5 ± 0.8		
HSV histogram	65.2 ± 1.2		

In the color based path analysis, the number of regions and the number of bins for computing HSV histogram had the values that gave the best result on image classification task, on this 8-category dataset, in preliminary tests. Similar to the edge based path, each data split contained 100 randomly selected training images and 50 test images for each category.

The best classification accuracy was achieved with the average of 144 regions (divided into 12 horizontally and 12 vertically) and 5 bins for H, S and V (total 15 bins). These parameters were used in subsequent evaluations, in addition to the parameters of C2 feature's 1600 prototypes extracted from training images (2 prototypes per image) and the C1 histogram's 5 regions and 25 bins which were decided based on the previous evaluation (see sec. 3.1).

For the full method evaluation, we chose the same test conditions as in previously published methods that have demonstrated state-of-the-art accuracy (87.8% [7], 83.7% [4]). Each data split contains 100 randomly selected training images and the rest of the images in each category are used as test images. Five independent data splits are analyzed. For further comparison, we evaluated one more method ('Stack') , in which all first stage vectors are simply concatenated and classified by an SVM (there is no second stage classifier).

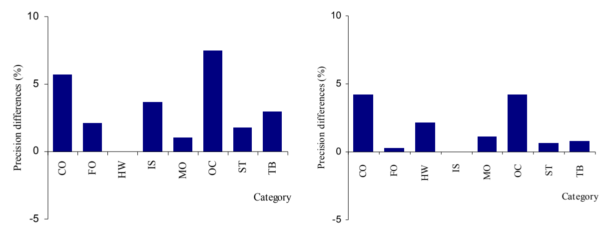
Table 2 shows the result of classifications with either the full feature method (i.e. C2 feature, C1 histogram and HSV), ‘C2 feature + C1 histogram’ (C2 feature and C1 histogram) or ‘C2 feature + HSV histogram’ (C2 feature and C1 histogram). The best result is obtained by the full feature with ‘Summation’ as the second stage classifier. In every case, the second stage classifier successfully improves the accuracy given by C2 feature. The parameters of this model, such as number of prototypes, histogram bins and regions, can be optimized beyond what is described in this evaluation. However it is clear that our full feature method is an improvement with respect to using shape based classification alone. More significantly, we demonstrated that the full feature method provides slightly superior performance (88.6% accuracy) relative to the previous state-of-the-art method (87.8% accuracy).

An interesting aspect of this evaluation is that even though HSV histogram by itself is not enough for accurate classification (65.2%), the color based path boosts the accuracy of the C2 feature, comparing ‘C2 feature + HSV histogram’ of 87.0% accuracy to C2 feature alone of 84.1% under the ‘Summation’ (see Table 2). Similarly, the full feature method is able to achieve 88.6% accuracy while ‘C2 feature + C1 histogram’ is 86.9% accurate without the contribution of the color.

Figure 4 (a) shows the detail of improvement in precision of each category. Each bar in the graph shows the difference of precisions: between ‘C2 feature + HSV histogram’ and only C2 feature on the left graph and between the full feature method and ‘C2 feature + C1 histogram’ on the right. None of categories suffer from a decrease in precision. Figure 4 (b) illustrates the example of the most significant improvement; ‘open-country’ increased by 7.5% with addition of HSV to C2 feature. ‘open-country’ is most often confused with ‘coast’ by C2 feature alone. Figure 4 (b) includes examples of images that are incorrectly classified as ‘coast’ by C2 feature but correctly as ‘open-country’ by ‘C2 feature + HSV histogram’.

4. Conclusion

We have described a biologically inspired method for image classification that utilizes three image properties which are shape based, edge based and color based. The method consists of two stages of image analysis. In the first stage, each image property is independently used for classification. In the second stage, a single classifier assigns the image category based on the probability distributions of the first stage classifier outputs. We demonstrate that our method can achieve a high accuracy which is comparable to the best reported methods on color scene image dataset. The better classification results obtained by adding edge or color based properties to shape based properties suggest that low level image features based on edges and color can boost classification accuracy when combined with higher level, shape-based, properties.



(a) The improvement of precisions on each category by adding HSV histogram. Left: added to C2 feature. Right: added to C2 feature and C1 histogram. (CO: coast, FO: forest, HW: Highway, IS: inside-city, MO: mountain, OC: open-country, ST: street, TB: tall building)

(b) Example images of ‘open-country’ that are correctly classified by ‘C2f + HSV’, while classified ‘coast’ incorrectly by C2 feature.

Figure 4. Improvement by HSV histogram on color image classification.

Appendix.

A-1. Reduction of Color information

In our study, we demonstrated that including color analysis improves the accuracy of image classification. Although color information in a particular scene is generally varied (e.g. the brightness of a coast scene on a sunny day is much different from the brightness on a cloudy day) and color descriptors are not able to classify scenes very well by themselves (see Table.2), the addition of color information helps C2 and C1 features classify scene images. This suggests that there might be a typical range of color descriptor values for semantic categories in a scene image data set, such as MIT Color NS, which is categorized by human subjects. Taking into account potential bias in the variety of color information in scene images, we supposed that all of the HSV descriptors, H (hue), S (saturation) and V (brightness) information, could be reduced effectively.

In this section, we investigate whether all HSV color descriptors are required to help the C2 feature and C1 histogram improve the accuracy of scene classification, or if a part of HSV is enough. We determined the accuracy that the full feature method can achieve on classifying the MIT Color NS dataset, with either all of hue, saturation and brightness histograms (HSV), with hue and saturation (HS) and with hue alone (H). The examination settings and data splits are as described in section 3.2. In the previous evaluation, the number of histogram bins for either of H, S or V in each region is 5, and the same number of bins is chosen in this evaluation. As the number of regions is 144, HSV has a total of 2160 histogram bins for an image (15 bins in a region), HS has 1440 (10 in a region) and H has 720 (5 in a region).

Table A-1 shows the result of the limited color information analysis, one with all of HSV, another with HS and the other with H alone. As the second stage classifier, a trained second SVM strategy is utilized in addition to the previously mentioned ‘Summation’. The accuracy with ‘Summation’ stays almost the same even when the HSV information is reduced from HSV (88.6%) to HS (88.6%) or H (88.4%), however if all color information is removed, accuracy decreases (86.9%). On the other hand, the accuracy of classification by only color information decreases when available color information is limited (from 65.2% of HSV to 61.6% of H alone).

Using the SVM as the second stage classifier does not improve the accuracy significantly but does provide higher accuracy than ‘Summation’ on HSV (88.9%) and H (88.7%). The results of SVM second stage classifier comparing HSV and H are shown in Figure A-1 as confusion maps. The similarity between the two maps suggests the potential for including only H color information in the full feature method. The result of the method with only H, suggests that this classification scheme is more robust than one with HSV, as smaller number of features is required while maintaining accuracy.

Table A-1. The results of color image classification with various aspects of color information.

	HSV	HS	H	No color
Summation	88.6 ± 0.2	88.6 ± 0.2	88.4 ± 0.2	86.9 ± 0.4
SVM	88.9 ± 0.3	88.1 ± 0.7	88.7 ± 0.6	87.0 ± 0.4
Only Color	65.2 ± 1.2	61.9 ± 1.0	61.6 ± 1.1	-

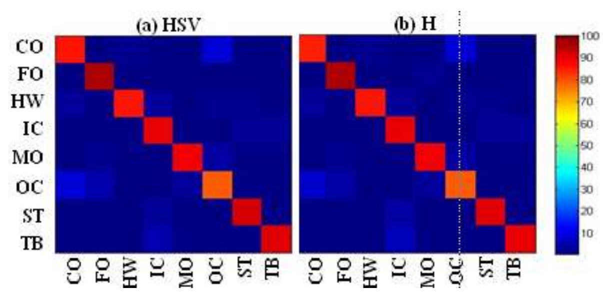


Figure A-1. The confusion matrixes of the full feature method analysis with either all of HSV (left) or H alone (right).

Acknowledgments

We thank Tomaso Poggio, Thomas Serra and Sharat Chikkerur for valuable suggestions and fruitful discussions. We also thank Aude Oliva, Antonio Torralba, Li Fei-Fei and Pietro Perona for providing the image data sets.

References

[1] M. Szummer and R. Picard. Indoor-outdoor image classification. *In IEEE Int. Workshop on Content-based Access of Image and Video Databases*, 1998.

[2] A. Vailaya, M. Figueiredo, A. K. Jain, and H.J. Zhang. Content-based hierarchical classification of vacation images. *In IEEE Conf. on Multimedia Computing and Systems,* 1:518-523, 1999.

[3] M. Pietikainen, T. Maenpaa and J. Viertola. Color Texture Classification with Color Histogram and Local Binary Patterns, *Proc. 2nd International Workshop on Texture Analysis and Synthesis*, June, 2002.

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

[5] L. Fei-Fei and P. Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. *IEEE Computer Vision and Pattern Recognition*, 2:524-531, 2005

[6] J. Vogel and B. Schiele. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*. 72(2): 133-157 April, 2007

[7] A. Bosch, A. Zisserman and X. Muoz. Scene classification using a hybrid generative/discriminative approach. *IEEE Trans. Pattern Analysis and Machine Intell.*, 30(04):712–727, 2008.

[8] S. Lazebnik, C. Schmid and J. Ponce, “Beyond bags features: spatial pyramid matching for recognizing natural scene categories”, *In IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2006.

[9] S. Thorpe, D. Fize and C. Marlot. Speed of processing in the human visual system. *Nature* 381:520–22, 1996.

[10] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Machine Intell.,* 29(3):411–426, 2007.

[11] V. Goffaux, C. Jacques, A. Mouraux, A. Oliva, P.G. Schyns and B. Rossion. Diagnostic colors contribute to early stages of scene categorization: behavioral and neurophysiological evidences. *Visual Cognition*, 12:878-892, 2005

[12] E. Deyoe and D.C. Van Essen. Concurrent Processing Streams in Monkey Visual Cortex. *Trends Neuroscience,* 11:219-226, 1998.

[13] D. Hubel and T. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.*148:574-591, 1959.

[14] I. Ivanov, B. Heisele and T. Serre. Using component features for face recognition. Proc. of the 6th International Conf. on Automatic Face and Gesture Recognition, 421-426, 2004

[15] M. M. Gorkani and R.W. Picard. Texture orientation for sorting photos “at a glance”. *In IEEE Intl. Conf. on Pattern Recognition,*1:459–464, 1994.

[16] B. Heisele, T. Serre and T. Poggio. A Component-based Framework for Face Detection and Identification. *International Journal of Computer Vision*, 74(2):167-181, 2007

[17] CBCL, http://cbcl.mit.edu/software-datasets /standardmodel/index.html

[18] M. Riesenhuber and T. Poggio. Hierarchical Models of Object Recognition in Cortex, *In Nature Neuroscience*, 2:1019-1025, 1999.