Beyond Shape: Incorporating Color Invariance into a Biologically Inspired Feedforward Model of Category Recognition

Jun Zhang*
Lab. of Image Information Processing
Hefei University of Technology
P.O. Box 98
Hefei, Anhui 230009
zhangjun1126@gmail.com

Zhao Xie
Lab. of Image Information Processing
Hefei University of Technology
P.O. Box 98
Hefei, Anhui 230009
xiezhao1980@126.com

Jun Gao
Lab. of Image Information Processing
Hefei University of Technology
P.O. Box 98
Hefei, Anhui 230009
gaojun@hfut.edu.cn

Kewei Wu
Lab.of Image Information Processing
Hefei University of Technology
P.O. Box 98
Hefei, Anhui 230009
wukewei1984@163.com

ABSTRACT

Being lack of theoretical support from biological cues in computer vision, current computational and learning approaches of object categorization mostly aim at better performances neglecting analysis on framework in human brain for visual information processing materially which cause little-marginal improvement and more complexity. Focusing on the uncertainty of color mechanism in visual cortex and motivating from biological issues on shape information, we present the model incorporating color invariant descriptors and plausible shape feature biologically to formulate the robust representation of each category with only simple *SVM* classifier to achieve the amazing performance. Our model has the characteristics of illumination, scale, position, orientation, viewpoint invariance, and competitive with current algorithms on only a few training examples from several data sets, including *Caltech 101* and *GRAZ* for category recognition. Also, experimental results show the robustness when challenged by noisy or blurred images.

Keywords

Color invariance, Biological-inspired, Object categorization, Shape description

*Corresponding author

ICVGIP '10, December 12-15, 2010, Chennai, India

1. INTRODUCTION

Human visual system can categorize objects rapidly and effortlessly despite the complexity and objective ambiguities of natural images. Despite the ease with which we see, visual categorization is an extremely difficult task for computers and is indeed widely acknowledged as a very difficult computational problem. The main issues lie in the variability of objects, such as scale, rotation, illumination, position, and occlusion. In previous work, it have been shown that scanning window strategy works relatively well for the recognition of objects with scale and position invariance [1]. Object categorization involves signal processing in neural circuit from a neuroscience perspective. If transmission and processing mechanism of visual information in the human brain were known, we would avoid designing unnecessary complex models (or algorithms). Much of the recent work on object recognition in computer vision lack physiological or psychological evidence.

According to the principle of 'what' pathway (i.e. ventral stream) in the human visual cortex [2], M. Riesenhuber and T. Poggio [3] extended Hubel and Wiese's hierarchical structure of visual cortex [4] and proposed the first feedforward model, called *'HMAX'*, mimicking visual information processing. Their model selects input with the strongest response by MAX-like operation, which is in fact a window analysis method from a computational perspective. Object recognition task was explained and supported by neuroscience data in their work. One drawback to this model is that they only learn and recognize computer-generated images of 3d wire frame (paperclip-like) stimuli, faces, and rendered cats and dogs. On the base of *'HMAX'* model, T. Serre et al [5, 6] created a biologically inspired feedforward framework (now frequently called the *'Standard Model'*) that mimics the first 150 milliseconds of the visual process in the brain and produces numerous shape-tuned units that are invariant to orientation, position and size during learning

stage. This is a biologically inspired model comparable to the state of the art in computer vision when applying into real-world scenes with high recognition rate on *Caltech 101* database.

Further, although the majority of images are recorded in color format in daily life, the explicit incorporation of color cues into object recognition systems has been largely ignored [7], only focusing on shape information to detect and extract features. Color values collected by sensors change greatly owing to variability in the outside world, which makes the color information hard to describe.

It is difficult to create neuron models of color information because of insufficient experiments. Studies on neuroscience have suggested that cells respond to colored stimuli more strongly than colorless one in Inferior Temporal (IT) and extrastriate visual areas V4 of the visual cortex. Other evidences indicate opponent color (i.e. red vs. green, blue vs. yellow) captures signals from cone cells and transmits them in the retina, Lateral Geniculate Nucleus (LGN) and visual cortex [8]. Psychophysical research has revealed that color, shape, depth and motion, as the subsystems of visual system, are interacted with each other in human perception. It was found that significant amount of visual processing is dedicated to the operation of color information [9]. Until now, how color works for object categorization has not arrived at a consistent conclusion.

In the case of feature extraction, full invariance is necessary. This naturally leads to our hypothesis that general object categorization is implemented not only by shape properties, but also by color cues. Our aim, nevertheless, is to design a computational model by taking into account some of neuroscience evidences present throughout cortex. In this paper, the model is designed according to the following criteria: (1) Features must target the orientation, scale, translation, photometric and geometry variations needed for applications. (2) Features must be robust against noise or blur, and should not contain instabilities. (3) Features must complement each other for occlusion.

To meet these criteria, we build on earlier work [5, 6] to extend the biologically inspired feedforward model and propose that color facilities object categorization in complex scenes or under variations. We propose a model to combine the biologically plausible shape features with color invariant descriptors.

The paper is organized as follows. In section 2, the biological mechanism of visual information in 'what' pathway is discussed. In section 3, an improved model incorporated with color invariant descriptors is proposed, considering reflection nature of color in physical optics. Section 4, provides several experiments and section 5 contains the concluding remarks.

2. BIOLOGICAL MECHANISM OF VISUAL INFORMATION IN 'WHAT' PATHWAY

In somewhat of an oversimplification, visual information in the brain transmits along two parallel and concurrent streams: the ventral ('what') stream and the dorsal ('where') stream. 'What' pathway processes visual shape, color, and texture appearance and is largely responsible for object categorization. Beginning at the retina, visual information arrives at Primary Visual Cortex (i.e. V1) through LGN, and then transmits to V2, V4 and IT. The receptive field sizes

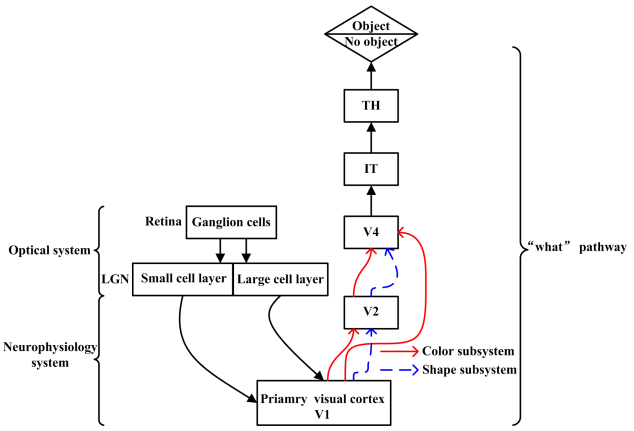

Figure 1: The processing and transmission of shape and color cues in the 'what' pathway of visual system.

and the position and scale invariance of the neural detectors increase along the stream [2, 10]. The receptive field size and preferred stimuli are shown in Table 1. Therefore, neuron mechanism can be considered as a feedforward hierarchical procedure in a low-high and bottom-up manner.

When light reflected by an object hits our photoreceptors of the retina, the decomposition of wavelength energy captured by retina sensory cells (i.e. cone cells), electrical impulses are created by retinal ganglion cells and sent out to other parts of the brain, that is V1, V2 and V4 of visual cortex along the 'what' pathway, which are important for the processing and perception of color [10]. Figure 1 shows information transmitted to LGN through retina is regarded as a human optics system, while information transmission in the 'what' pathway is viewed as a neurophysiology system. Information passes to small cell layer and large cell layer of LGN with the help of neuron cells as soon as it arrives at the retina. Then different visual cues, like color and shape, separate into two different subsystems in V1. The red solid line indicates information processing in the color subsystem, from V1 to V2 or V4, the blue dashed line shows shape cues processing in the visual cortex, from V1, V2 to V4.

3. OUR MODEL

Color responses strongly in IT (V4), and where opponent color response exists in the entire visual cortex. Thus shape-based model incorporated with color cues may increase the performance on object categorization significantly. Algorithmically, this computation can be performed borrowing color story from physical optics [11]. We combine color descriptions invariant to photometric and geometric robustness with biologically plausible shape features.

Here we use the framework of object recognition from T. Serre et al [5, 6]. In this framework, the integrated features ($\boldsymbol{F}$), which will be used to represent objects, are computed by combining two sources of information: shape features ($\boldsymbol{S}$), and a model of color descriptors ($\boldsymbol{C}$).

$$\boldsymbol{F} = [\boldsymbol{S}, \alpha \boldsymbol{C}] \tag{1}$$

Where the coefficient α acts like a weight parameter. The parameter α is set by sequentially searching for the best α on a validation set. The optimization was achieved by

Table 1: Preferred stimuli and receptive field size in the hierarchy of 'what' pathway

visual cortex	V1	V4	IT
preferred stimuli	oriented bars, edge, retina position	angle, color, shape, texture	objects such as face
receptive field size	~1.5°	~4°	~26°

using *Caltech 5* as the target object (see details in the first experiment). However, we found this parameter had a small effect when the target object was changed. The parameter α was then fixed for all the experiments.

Figure 2 illustrates an overview of the data flow diagram. The right part of the scheme, showed as yellow elliptic boxes, corresponds to the feedforward shape model. Details about the model can be found elsewhere [5, 6]. In the following we provide a short description of its implementation.

Gabor filter, similar as reflective field properties of simple cells, provides selectivity to specific frequency and orientation, that is, once the image is filtered, features that correspond to specific frequency and orientation can be obtained. Thus S1 cells response at scale $\boldsymbol{s}$ and direction $\boldsymbol{d}$ is computed using Gabor filters with 4 orientations $(0, \pi/4, -\pi/4, \pi/2)$, 16 different scales (from 7×7 to 37×37 pixels in steps of two pixels, forming 8 scale bands), thus leading to 64 different filter responses. Next, the C1 cell responses are computed by MAX-like operation. That is, we take a max over the two scales within the same orientation which shows some tolerance to shift and size. The results are 32 responded images (8 scales×4 orientations). In the S2 layer, we firstly extract a set of prototype features for the S2 units at the level of the C1 layer across all four orientations, i.e., a patch of size n×n×4, n=4, 8, 12 and 16. The response of an individual S2 unit is given by matching (similar to Gaussian-like tuning.) on the Euclidean distance between a new input and a stored prototype. Thus we obtain S2 maps on 8 bands. Our final set of shift- and scale-invariant C2 responses (a vector $\boldsymbol{S}$ of 1,000 values) is computed by taking a global maximum over all scales and positions for each S2 type.

The left part of this scheme in green elliptic boxes shows the extraction of color descriptors. Although color cues can not be built by neuron models, Dichromatic Reflection (DRF) model [11] explains how the image formation and photometric changes (such as shadow, light source, specularities) influence RGB values of images. On the basis of DRF model, color models are discussed containing invariance in the following.

We assume that objects consist of different materials (including e.g. papers and plastics), the RGB values obtained by sensors with spectral sensitivities $f^C(\lambda)$ are:

$$\vec{C}(x) = [R, G, B]^T = m_b(x)\vec{C}_b + m_i(x)\vec{C}_i + \vec{C}_a \tag{2}$$

$$\begin{aligned} C(x) = m_b(x) \int_\lambda e^C(\lambda) f^C(\lambda) c_b(\lambda) d\lambda + \\ m_i(x) \int_\lambda e^C(\lambda) f^C(\lambda) c_i(\lambda) d\lambda + \int_\lambda a^C(\lambda) f^C(\lambda) d\lambda \end{aligned} \tag{3}$$

For $C \in R, G, B$, and where $\vec{C}_i$ is color of specular reflectance light that is immediately reflected at the surface, causing highlights. $\vec{C}_b$ and $\vec{C}_a$ are the body reflectance and diffuse light caused by reflectance from all the directions. The geometric terms of the reflectance $m(x)$ are the geometric dependencies on the viewing angle, light source direction and surface orientation. Suppose body reflectance is

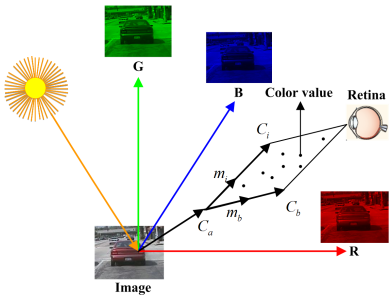

Figure 3: Schematic diagram for color information entering into human retina when sunlight shines object.

Neutral Interface Reflection (NIR) model, meaning Fresnel reflectance $c_i(\lambda) = c_i$. Furthermore, $e^c(\lambda)$ is a single light source, $a^c(\lambda)$ is the diffuse light, where λ is the wavelength. When light shines upon one object, the RGB-colors formed by reflectance light superposition on the surface of the object are captured by human retina, see Fig. 3.

For simplicity, we assume white illumination, i.e. all light sources in the scenes have constants: $e^C(\lambda) = e$, and assume that the following holds: $f^C(\lambda) = \delta(\lambda - \lambda_C)$. With these assumptions, we have the following equation for the sensor values from an object under white light:

$$C(x) = m_b(x)c_b^C(\lambda)e + m_i(x)c_i e + a^C(\lambda) \tag{4}$$

Equation 4 is reflectance function of objects in scenes.

3.1 Photometric invariance

Opponent color theory assumes retina has three visual components, corresponding to retinal ganglion cells, which transmit in the LGN and visual cortex. The three components are defined as

$$O_1 = \frac{G+B+R}{\sqrt{3}}, O_2 = \frac{R-G}{\sqrt{2}}, O_3 = \frac{G+R-2B}{\sqrt{6}} \tag{5}$$

Where O_1, O_2, and O_3 are white-black, red-green, and blue-yellow components, respectively. Here, we only consider (O_2, O_3), because color changes don't affect O_1 , and the case for which there is no diffuse illuminant present $(a^C(\lambda) = 0)$. By substituting eq. 5 in the eq. 4 we obtain

$$O_2 = \frac{m_b(x)e[c_b^R(\lambda) - c_b^G(\lambda)]}{\sqrt{2}} \tag{6}$$

$$O_3 = \frac{m_b(x)e[c_b^G(\lambda) + c_b^R(\lambda) - 2c_b^B(\lambda)]}{\sqrt{6}} \tag{7}$$

Obviously, the above equation is invariant for $m_i(x)$, that is, in the absence of diffuse light, invariance with respect to the highlight can be obtained.

Since hue can be treated as the angle of colors in the

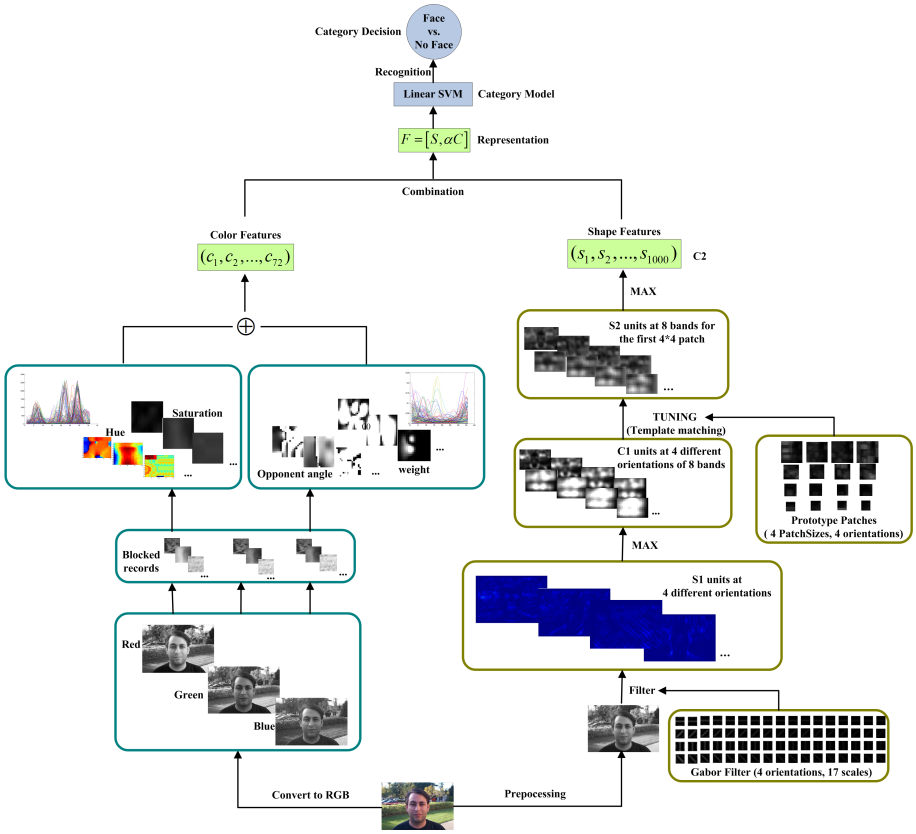

Figure 2: The extended model incorporated with color invariant descriptors.

(O_2, O_3) space [12], that is,

$$\begin{aligned} \boldsymbol{H}(R,G,B) &= \arctan(\frac{O_2}{O_3}) \\ &= \arctan(\frac{\sqrt{3}(c_b^R(\lambda) - c_b^G(\lambda)}{c_b^R(\lambda) + c_b^G(\lambda) - 2c_b^B(\lambda)}) \end{aligned} \tag{8}$$

Invariance with respect to both the lighting geometry $m_b(x)$ and specularity $m_i(x)$ is obtained by hue.

We consider hue ($\boldsymbol{H}$) as a color invariant descriptor. To suppress the effect of noise for unstable color invariant values, an effective object representation is on the basis of histogram [13]. The predicted uncertainty in $\boldsymbol{H}$ is given by

$$\begin{aligned} \sigma_H &= \sqrt{(\frac{\partial H}{\partial O_2}\sigma_{o_2})^2 + (\frac{\partial H}{\partial O_3}\sigma_{o_3})^2} \\ &= \sqrt{\frac{1}{O_2^2 + O_3^2}} \end{aligned} \tag{9}$$

Where σ is the standard deviation of opponent colors, and we set $\sigma = 1$.

Opponent color is known to be sensitive to saturation that indicates the color amplitude. The smaller the saturation is the more uncertain the hue estimation [12]:

$$\begin{aligned} S(R,G,B) &= \sqrt{O_2^2 + O_3^2} = \frac{1}{\partial H(R,G,B)} \\ &= (\frac{2}{3}(m_b e)^2[(c_b^R)^2 + (c_b^G)^2 + (c_b^B)^2 \\ &- c_b^R c_b^G - c_b^R c_b^B - c_b^G c_b^B])^{\frac{1}{2}} \end{aligned} \tag{10}$$

Thus we can weight H_{hist} by its saturation.

3.2 Geometric invariance

Suppose that the color of an edge can locally be modeled as a smoothed step edge. It is now straightforward to prove that the angles between the derivative-based color channels are invariant to this smoothing and are only dependent on the edge amplitude [14]. We introduce a new invariant, called 'color angle', to implement the geometric invariance. The derivatives of the colors

$$C_x(x) = m_{b_x}(x)c_b^C(\lambda)e + m_b(x)c_{b_x}^C(\lambda)e + m_{i_x}(x)c_i e \tag{11}$$

are invariant to diffuse light $a^C(\lambda)$. If we subsequently consider specular reflection, we obtain the derivatives of the opponent colors

$$\begin{aligned} O_{2_x} &= R_x - G_x \\ &= \frac{1}{\sqrt{2}}m_{b_x}(x)e[c_b^R(\lambda) - c_b^G(\lambda)] + m_b(x)e[c_{b_x}^R(\lambda) - c_{b_x}^G(\lambda)] \end{aligned} \tag{12}$$

$$\begin{aligned} O_{3_x} &= G_x + R_x - 2B_x \\ &= \frac{1}{\sqrt{6}}m_{b_x}(x)e[c_b^R(\lambda) + c_b^G(\lambda) - 2c_b^B(\lambda)] \\ &+ m_b(x)e[c_{b_x}^R(\lambda) + c_{b_x}^G(\lambda) - c_{b_x}^B(\lambda)] \end{aligned} \tag{13}$$

This can be proven to be invariant with respect to specular variations, similarly as in Eq. 6 and Eq. 7.

We can now add the geometrical invariance to the photometrical invariant derivatives, this leads to the opponent angle:

$$\theta_x = \arctan(\frac{O_{2_x}}{O_{3_x}}) \tag{14}$$

Similarly as for the hue we apply an error analysis to the color angle equations of Eq. 14, which yields the following results of predicted uncertainty in θ_x

$$\sigma_\theta = \sqrt{(\frac{\partial H}{\partial O_{2_x}}\sigma_{O_{2_x}})^2 + (\frac{\partial H}{\partial O_{3_x}}\sigma_{O_{3_x}})^2} = \sqrt{\frac{1}{O_{2_x}^2 + O_{3_x}^2}} \tag{15}$$

Hence, we will use σ_θ as the weight for the opponent angle histogram $\theta_{x_{hist}}$ when converting it to a local color histogram.

So far, we obtain color descriptors by concatenating a photometric invariance descriptor H_{hist} to the geometric invariance descriptor $\theta_{x_{hist}}$, according to

$$C = [H_{hist}, \theta_{x_{hist}}] \tag{16}$$

In Table 2 an overview of the invariants in our model is given.

4. EXPERIMENTS AND DISCUSSION

We evaluate our model by six experiments, each of which focuses on different performance: (1) Performance obtained with different weights α, (2) binary classification, (3) the influence of the number of training examples on the performance of the model, (4) blurred or noised robustness, (5) multiclass classification, and (6) the performance on the complex image database (including occlusion, intra-class variations, etc).

4.1 Image data sets

Each image set was randomly split into two disjoint sets of training and test images. In our experiments, no single object was present in both sets.

Caltech 5: We considered five of the databases, i.e., Faces, Motorbikes, Cars (rear), and Airplanes data sets [15], as well as Leaves data set [16] (see typical examples in Fig. 4(a)) for the former five experiments. All images were normalized to 140 pixels in height (width was rescaled accordingly so that the image aspect ratio was preserved) and converted to gray scale before processing.

Caltech 101: This datasets contain 101 objects plus a background category (used as the negative set) [15], which have been a standard database owing to objects in it are involved in several circumstances in the real world, such as light, view, orientation changes. All images were preprocessed as *Caltech 5*.

GRAZ: The dataset is very challenging: it contains a large variety of objects under wide pose and extreme illumination conditions [16]. Here we used cars, person and bikes (see typical examples in Fig. 4(b)).

4.2 Results

(1) Weight optimization

In order to validate our weight α, we tested the performance, under the classification task described in Section 3. We learned models with different α ($\alpha \in \{0, 1\}$) for *Caltech 5* dataset. For algorithm effectiveness, we only took 5 positive training examples, 20 negative images for training and

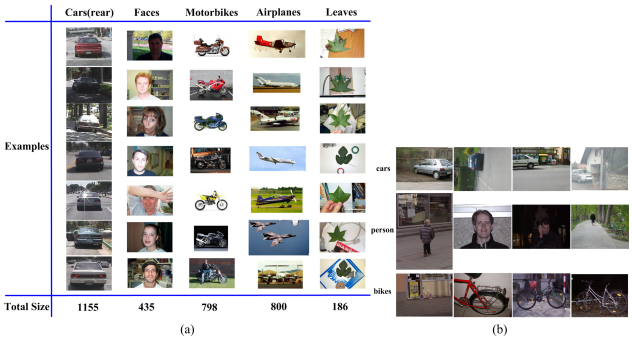

Figure 4: Examples of *Caltech 5*(a) and *GRAZ* database(b).

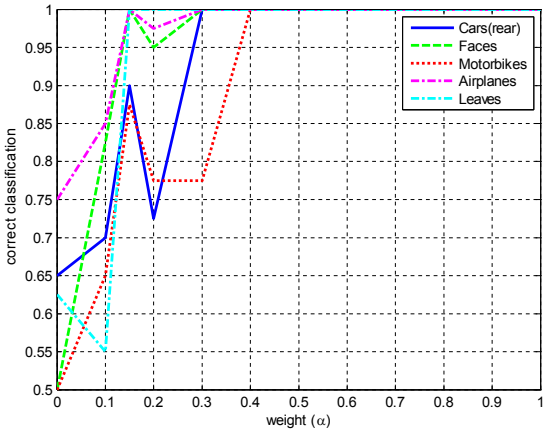

Figure 5: Performance for all 101-object categories for different weights α.

40 images for testing (half positive, half negative). Figure 5 shows the results that summarize the success of the overall features in the increase of the parameter α. Upon inspection of the different α selected across model, it is necessary to explore the range of parameters [0.5, 1] with $\alpha = 0.6$ in order to guaranty that perfect performance can be achieved given the parameter α within our experiments. A small value for α ($\alpha < 0.5$) has the effect of downweighting the importance of color with respect to shape information. It has to be noted that, when no color information ($\alpha = 0$) is taken into account, the performance declined significantly by 25% at least for airplane category.

This suggests that although regions that have different shape properties than their neighborhood are often considered salient (more informative) and attract attention, color features with photometric and geometric invariants, as a complementary cue, can boost categorization performance.

(2) Binary classification

This experiment was carried out as follows: the model was trained on the first set and tested on the second one. The results reported were averaged over 3 independent runs. The performance figures quoted are ROC equal error rates. As Fig. 6 shows, surprisingly, the performances of our model on all five categories are almost perfect. Amongst the datasets, only the faces involve small changes. Table 3 summarizes the

Table 2: Properties of descriptors in each stage ('+' indicates the feature is insensitive to the property, otherwise '-' indicates sensitive)

Features	Orientation	Scale	Translation	View	Highlight	Illumination geometry	Diffuse light
H	-	-	-	-	+	+	-
θ	+	-	+	+	+	-	+
C1	-	+	+	-	-	-	-
C2	+	+	+	-	-	-	-

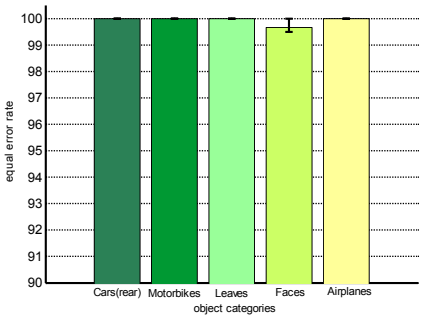

Figure 6: Equal error rate on the *Caltech 5* database.

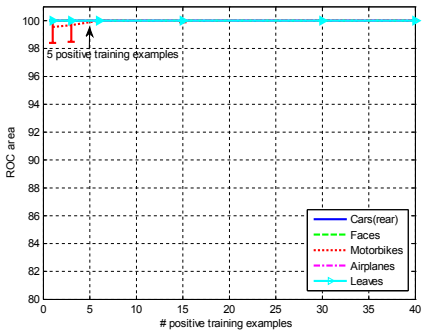

Figure 7: Performance obtained with improved model and different numbers of training examples.

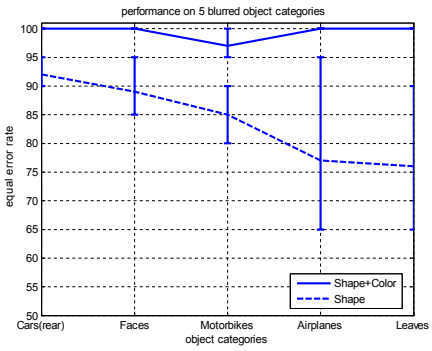

Figure 8: Equal error rate between improved model and shape model for blurred images.

performance of our model compared with other published results from benchmark systems.

This experiment verifies that color cues facilitate object categorization. But it should be noted that the running time is long owing to the training set size. In the next experiment it will be shown how well our model can arrive given less training examples.

(3) Different numbers of training examples

For each of five object category, the model was trained with 1, 3, 6, 15, 30 and 40 positive training examples and 50 negative training examples from the background class. From the remaining images, we extracted 50 images from the positive and 50 images from the negative set to test the system's performance (each result is an average of 5 different random splits). The performance measure reported is the area under the ROC (AUC). Figure 7 illustrates the recognition performance for the different training examples. A strong performance gain is observed for the majority of the 5 categories, even when only a few training examples are used. At Training Number = 3, the model achieves the performance near 100%. Looking back to the last two experiments, it is no wonder that they have succeeded. At Training Number = 5 that our model achieves an average performance of 100%, which is shown by the black arrow. With less than 5 training examples, the model achieves recognition performance comparable to [5, 6] using 40 training examples, that is, the extended model clearly show a big advantage over shape model when training number is small.

(4) Robust to blurred or noisy images

Actually, the captured images present with blur or noise probably due to inevitable limitation of imaging sensor. We here describe an experiment that suggests that it is possible to perform robust object categorization with our model learned from low quality images. It is known from the last experiment that the model gets nearly 100% correct using 5 training examples. Thus the *Caltech 5* dataset used here contains 5 positive and 20 negative training examples and 40 test examples (half positive, half negative). Classification scores for our model were averaged over 5 runs. Figure 8 shows the comparison of equal error rate across all 5 categories between our model (called Shape + Color model) and T. Serre's shape model. Overall, for 4-object categories tested, results obtained with our model achieve perfect performance. For the worst case of motorbike category, our model gets 97.5% correct (equal error rate = 0.95). The performance of shape model degrades by 35% for leave and airplane categories.

It is interesting to investigate why objects can not be recognized correctly without color cues. We take faces, cars (rear) and leaves for examples. Figure 9 shows missed examples of shape model and their color invariant histograms. The top and second lines illustrate original and blurred images we created, and their corresponding color descriptors. It can be seen that blur don't influence color information, so color may play a positive role in object recognition. The bot-

Table 3: Comparison between improved model and the state of the art

Datasets	Cars(rear)	Faces	Motorbikes	Airplanes	Leaves
Our model	100	100	100	100	100
Shape model [5, 6]	99.8	98.1	97.4	94.9	95.9
Constellation models	84.8 [17]	96.4 [17]	95.0 [17]	94.0 [17]	84.0 [18]

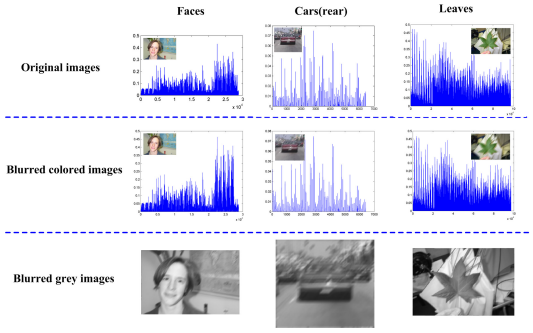

Figure 9: Some missed examples of shape model.

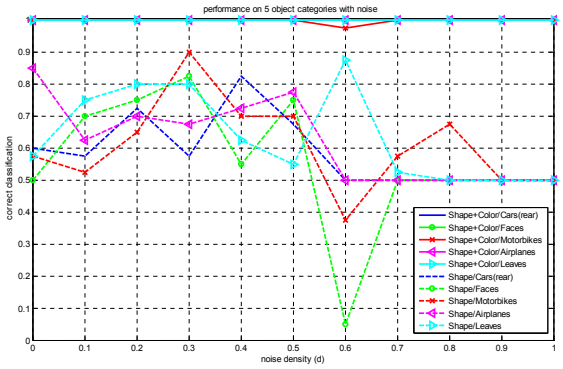

Figure 10: Correctness rate of two models with different noise densities.

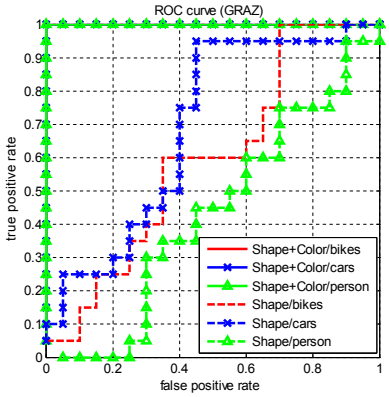

Figure 11: ROC curves for cars, persons and bikes with 5 training examples on *GRAZ* database.

tom line shows blurred grey images with less clearer shape, which is the reason for significant decline in performance without color cues.

Similarly, we reported results on binary classification on the *Caltech 5* database with different noise density (d). To conduct this experiment, we generated different testing sets of noise density 0, 0.1 $\cdots$ 0.9, and 1. Figure 10 compares the influence of noise on recognition performance of our model to that of shape model. It is seen that our model is markedly superior to those from shape model. The sharp difference of performance between these two models is 95%. Table 5 shows the detailed performance across the five datasets.

In general, under degradation, purely shape representations are not enough for accounting for the reliable object recognition performance when the object is presented in clutter. Our model is more robust to image degradation.

(5) Multiple classes case

We conducted initial experiments on the multiple classes case. For this task we used the *Caltech 101* dataset. We split each category into a training set of size 5 and a test set containing 20 images, similar to the experiments described earlier. The classifier is a simple multiclass linear SVM that applied the all-pairs method for multiple label classification and is trained on 102 labels. We obtained above 36% correct classification rate only using 5 training examples per class, comparable to [6] on 15 training examples.

(6) Object recognition in complex scenes

In order to test our model on a challenging real-world object recognition problem, we have built training and test data from the *GRAZ* data set, as shown in Fig. 4 (b). This database consists of wide internal variability in their appearance. For example, the object class car includes examples of many diverse models, at many poses, and in various types of occlusion and lighting, person appears at different locations, and the class of bikes includes bicycles as well as electric motor car. Our results are shown in Fig. 11 along with those of shape model. The performance could be improved significantly if the model was incorporated with color information. This is due to color features are important for light, viewpoint changes and occlusion in objects.

5. CONCLUSION

Biology research suggests that color plays an important role in the information processing of human visual cortex, especially for opponent color responding strongly to IT (V4). In this study we have shown that a biologically-based model with color cues, a color descriptor with photometric and geometric invariance in the opponent color space, strengthens the case for investigating biologically-motivated approaches to this problem, can compete with other state-of-the-art approaches to object categorization. The categorization results presented here convincingly exhibit excellent performance of our model: Interestingly, the approach was shown to be able to learn from a few examples and robust to the images with low quality, even for a variety of real-world object recognition tasks. The goal of our approach is to develop a model

Table 4: Correctness rate of improved model and shape model with different noise densities (d) (the left of '/' denotes our model, the right shape model)

Object category	d=0.5	d=0.6	d=0.7	d=0.8	d=0.9	d=1
Cars(rear)	1/0.6750	1/0.5000	1/0.5000	1/0.5000	1/0.5000	1/0.5000
Faces	1/0.7500	1/0.0500	1/0.5000	1/0.5000	1/0.5000	1/0.5000
Motorbikes	1/0.7000	0.9750/0.3750	1/0.5750	1/0.6750	1/0.5000	1/0.5000
Airplanes	1/0.7750	1/0.5000	1/0.5000	1/0.5000	1/0.5000	1/0.5000
Leaves	1/0.5500	1/0.8750	1/0.5250	1/0.5000	1/0.5000	1/0.5000

built on the visual cognition, not pursue complicated algorithms directly. Finally it is worth pointing out that there are also horizontal, within-area and feedback connections. It is likely that future work will study the role of feedback connections.

6. ACKNOWLEDGMENTS

We would like to thank our reviewers for helping to improve the clarity of the paper. This research was sponsored by grants from: the National Natural Science Foundation of China (No. 60875012 and No. 60905005), the National Research Foundation for the Doctoral Program of Higher Education of China.

7. REFERENCES

[1] H Schneiderman and T Kanade. Object detection using the statistics of parts. *International Journal of Computer Vision.*, 56(3):151–177, March 2004.

[2] S Kastner and L G. Ungerleider. Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23(1):315–341, March 2000.

[3] M Riesenhuber and T Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 23(2):1019–1025, March 1999.

[4] D H. Hubel and T N. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, October 1959.

[5] T Serre, L Wolf, and T Poggio. Object recognition with features inspired by visual cortex. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 994–1000. IEEE Computer Society, June 2005.

[6] T Serre, L Wolf, S Bileschi, and T Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):411–426, March 2007.

[7] J Weijer, T Gevers, and A Smeulders. Robust photometric invariant features from the color tensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(1):118–127, January 2006.

[8] S Borer and S Sĺźsstrunk. Opponent color space motivated by retinal processing. In *First European Conference on Color in Graphics, Imaging and Vision*, pages 187–189. IEEE Computer Society, June 2002.

[9] M S. Livingstone and D H. Hubel. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *The Journal of Neuroscience*, 7(1):3416–3468, November 1987.

[10] J Tanaka, D Weiskopf, and P Williams. The role of color in high-level vision. *Trends in Cognitive Sciences*, 5(1):211–215, April 2001.

[11] S A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10(4):210–218, February 1985.

[12] Th. Gevers, J. van de Weijer, and H. Stokman. *Color Feature Detection: An Overview*. CRC Press, Boca Raton, Florida, 2006.

[13] T Gevers and H Stokman. Robust histogram construction from color invariants for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(26):113–118, January 2001.

[14] Joost van de Weijer and Cordelia Schmid. Coloring local feature extraction. In *European Conference on Computer Vision*, pages 334–348. IEEE Computer Society, June 2006.

[15] F. 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 *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 18–32. Workshop on Generative-Model Based Vision, June 2004.

[16] A Opelt, A Pinz, M Fussenegger, and P Auer. Generic object recognition with boosting. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 28(3):416–431, March 2006.

[17] Fergus R. Perona P and Andrew Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 264–271. IEEE Computer Society, June 2003.

[18] M Weber, M Welling, and P Perona. Unsupervised learning of models for recognition. In *European Conference on Computer Vision*, pages 18–32. IEEE Computer Society, June 2000.