A New Biologically Inspired Color Image Descriptor

Jun Zhang[1,2], Youssef Barhomi[1], and Thomas Serre[1,⋆]

[1] Department of Cognitive Linguistic & Psychological Sciences
Institute for Brain Sciences
Brown University, Providence, RI 02912, USA
[2] School of Computer & Information
Hefei University of Technology, Hefei, Anhui 230009, China
zhangjun1126@gmail.com,
{youssef_barhomi,thomas_serre}@brown.edu

Abstract. We describe a novel framework for the joint processing of color and shape information in natural images. A hierarchical non-linear spatio-chromatic operator yields spatial and chromatic opponent channels, which mimics processing in the primate visual cortex. We extend two popular object recognition systems (i.e., the Hmax hierarchical model of visual processing and a sift-based bag-of-words approach) to incorporate color information along with shape information. We further use the framework in combination with the gist algorithm for scene categorization as well as the Berkeley segmentation algorithm. In all cases, the proposed approach is shown to outperform standard grayscale/shape-based descriptors as well as alternative color processing schemes on several datasets.

Keywords: image descriptor, color, Hmax, sift, bag-of-words, gist, object recognition, scene categorization, segmentation.

1 Introduction

Color constancy, which ensures that the perceived color of objects is tolerant to changes in lighting, is widely acknowledged to be a challenging computational problem. The primate visual system devotes impressive resources to the processing of color information. It has been suggested that color processing may provide a cue to scene parsing and figure-ground segmentation [1], and to the recognition of certain classes of objects (e.g., fruits), particularly under degraded viewing conditions [2]. How the visual system processes color information and achieves constancy remains, however, a matter of debate.

Two functional classes of color-sensitive neurons have been described [3]: Single-Opponent (SO) and Double-Opponent (DO) neurons. It has been suggested that SO cells are involved in the processing of surface information because

[⋆] Supplementary online material (including software and additional figures) available at http://serre-lab.clps.brown.edu/projects/color.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part V, LNCS 7576, pp. 312–324, 2012.

they exhibit strong selectivity for color opponency (e.g., red vs. green) and weak tuning for spatial opponency (i.e., orientation). DO cells, on the other hand, are thought to be involved in the extraction of boundary information and tend to be responsive to both color and spatial opponency (e.g., oriented red bar on a green background)[1]. DO cells may provide a biological basis for Land's retinex algorithm [4] and thus play a key role in color constancy.

In computer vision, there have been two main approaches to color processing for object recognition. One standard approach consists in applying popular shape-based image descriptors such as the SIFT descriptor on individual color channels, e.g., HSVSIFT [5], OpponentSIFT [6] and the related CSIFT descriptors [7] (based on a retina-like post-receptoral opponent color space). Another color descriptor, which is robust to photometric changes, uses a local color tensor [8]. More recently, a multi-spectral SIFT descriptor was proposed [9] whereby a 4D RGB+NIR color vector was used in combination with the SIFT descriptor to represent spatial and chromatic cues.

Another approach to color processing involves the concatenation of shape-based descriptors (computed from grayscale pixel intensities) with hue/color histograms [10,11]. In such approaches, color and shape information are thus processed separately. Computational methods have been proposed for the construction of improved color histograms [12]. Van de Weijer & Schmid described the HueSIFT descriptor for image classification [10], which is based on the concatenation of a hue histogram with grayscale SIFT image descriptors. In a recent study [6], color image descriptors were evaluated in terms of their robustness to common image transformations. While the relative ranking of the descriptors varied depending on the dataset, the OpponentSIFT was shown to be the best choice in the absence of any task-specific prior knowledge.

Here we describe a hierarchical model of color processing that tries to mimic the anatomy and physiology of the primate visual cortex. The recursive application of a new non-linear spatio-chromatic operator yields spatial and chromatic opponent channels that resemble processing by the SO and DO cells described above. As we will show, such a hierarchical processing outperforms approaches that simply concatenate color and shape cues extracted separately. Similarly, the proposed hierarchical scheme outperforms simpler shallow approaches such as the OpponentSIFT descriptor whereby gradient information is computed directly from post-receptoral opponent channels independently.

We extend the SIFT descriptor and the biologically inspired HMAX model [13] to include color cues for object recognition and report on their performance on several color datasets (i.e., the soccer team dataset [10] and the flower dataset [14]) as well as the PASCAL VOC 2007 challenge [15]. We further describe an extension of the popular GIST algorithm with an application to scene categorization [16] as well as an evaluation of the proposed pipeline for contour detection in natural images using the BSDS500 benchmark [17].

[1] Also most DO cells tend to respond to both chromatic and non-chromatc stimuli.

2 Spatio-chromatic Opponent Descriptors

Here we describe the processing pipeline necessary for the computation of the SO (surface) and DO (boundary) descriptors. Classical opponent theories of color vision have emphasized two main (equiluminant) chromatic axes: *Red-Green* (R-G) and *Yellow-Blue* (Y-B) [18] obtained from the combination of individual (R, G and B) color channels via antagonist center-surround receptive fields (RFs). We here further consider a *Red-Cyan* (R-C) channel (C channel obtained by further combining G and B channels) as well as a (luminance-based) *White-Black* (Wh-Bl) channel (Wh and Bl channels are obtained by combining the R, G and B channels together). Whether the R-C channel constitutes a separate color channel in its own right or instead corresponds to a noisy R-G channel (whereby blue cones would mingle with green cones because of imperfect wiring) remains a matter of controversy [19]. We found that the addition of this R-C channel consistently increases the performance of the proposed approach by about 2–5%.

Single-Opponent (SO) Descriptor: Processing starts with an RGB or LMS input image and comprises two stages. In stage I, four pairs of opponent color channels are first created using linear combinations of filtered color channels (see Fig. 1). The response of a model unit is obtained by considering the dot-product between an image patch $I(x, y, \lambda)$ and the spatio-chromatic sensitivity function given by:

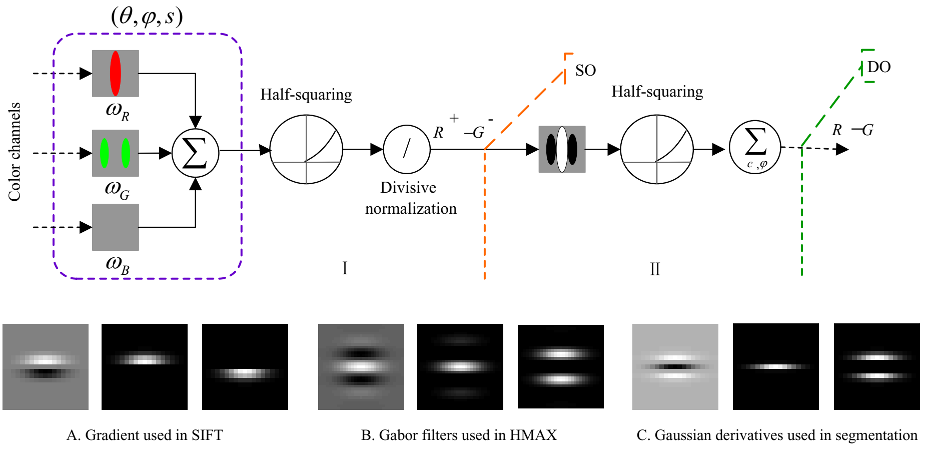
$$f(x, y, \lambda) = \omega_R R(\lambda) f_R(x, y) + \omega_G G(\lambda) f_G(x, y) + \omega_B B(\lambda) f_B(x, y), \qquad (1)$$

where $R(\lambda)$, $G(\lambda)$, and $B(\lambda)$ correspond to the spectral response functions (in practice we use the standard R, G, B components from color images for computer vision applications but more realistic spectral response functions including LMS could be used). $f_R(x, y)$, $f_G(x, y)$, and $f_B(x, y)$ correspond to the spatial sensitivity distributions for each individual color component. These are obtained by isolating the positive/negative subunits from linear oriented filters to form excitatory/inhibitory center or surround structures. We have used three kinds of filters depending on the application and benchmark used: gradient operator, Gabor filters, and Gaussian derivatives (see Fig. 1 and Sec. 3 for details).

The matrix Ω below contains the weight vectors used to combine the R, G, B channels into four pairs of opponent color channels (individual weight vectors $\boldsymbol{\omega} = (\omega_R, \omega_G, \omega_B)$ for each pair are stored as column vectors), s.t.:

$$\Omega = \begin{pmatrix} \pm 1/\sqrt{2} & \pm 2/\sqrt{6} & \pm 1/\sqrt{6} & \pm 1/\sqrt{3} \\ \mp 1/\sqrt{2} & \mp 1/\sqrt{6} & \pm 1/\sqrt{6} & \pm 1/\sqrt{3} \\ 0 & \mp 1/\sqrt{6} & \mp 2/\sqrt{6} & \pm 1/\sqrt{3} \end{pmatrix}$$

In the example shown in Fig. 1, $\boldsymbol{\omega} = (+1/\sqrt{2}, -1/\sqrt{2}, 0)$ such that the "+" (resp. "-") sign indicates an excitatory red center (reps. inhibitory green surround) component. This process can be thought of as a 3D convolution between a color image and a non-separable (spatio-chromatic opponent) operator. The corresponding RFs exhibit some selectivities for opponent colors and are typically weakly oriented.



A. Gradient used in SIFT

B. Gabor filters used in HMAX

C. Gaussian derivatives used in segmentation

Fig. 1. Proposed spatio-chromatic opponent image descriptors. **Top:** Individual R, G, B color channels are first convolved with a center and a surround filter at orientation θ, phase φ, and scale s. The corresponding color channels are further combined (see text for detail) and rectified by half-squaring followed by divisive normalization (I). This yields eight chromatic SO channels organized in four pairs (e.g., R^+–G^- shown here and R^-–G^+). In stage II, an oriented filter (with both excitatory and inhibitory subunits) is further applied on each SO channel, followed by half-squaring rectification and summation over phases and opponent pairs to yield four spatio-chromatic DO channels that are invariant to figure-ground reversal (e.g., R-G). **Bottom:** Original filters and corresponding subunits used. (**A**) Gradient used in SIFT [20], (**B**) Gabor filters used in HMAX [13] and (**C**) Gaussian derivatives used in segmentation algorithms [17] (additional filters used at multiple orientations, scales and phases not shown).

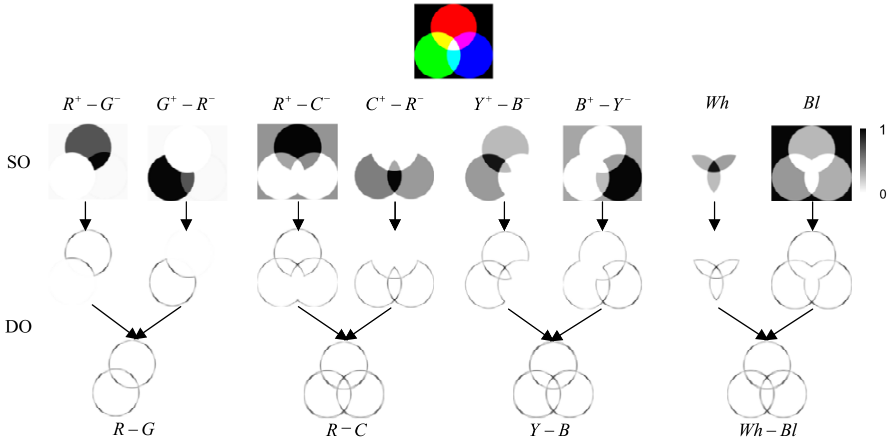


Fig. 2. Processing by the SO and DO color channels. Also note that color regions in the original image are not equiluminant and, as a result, the strength of the SO model unit responses varies across image locations.

Table 1. On the need for non-linear circuits: Recognition performance on the soccer team and Pascal voc 2007 datasets with and without rectification or divisive normalization stages for the SO (left) and DO (right) sift descriptors.

Methods	Soccer team	Pascal voc 2007
Full model	82.0 / 66.0	33.3 / 39.8
Without half-squaring	62.0 / 60.0	30.3 / 36.7
Without normalization	70.0 / 53.0	32.9 / 40.7

The response of the spatio-chromatic opponent operator is rectified (half-squaring) to maintain positive firing rates [21]. We further apply an extension of the divisive normalization circuit (originally proposed to model the contrast response of cells in the primary visual cortex [22]) to color processing. This step can be described by the following equation:

$$v(x, y, c) = \sqrt{\frac{k \times u(x, y, c)}{\sigma^2 + \sum u(x, y, c)}}, \tag{2}$$

where $u(x, y, c)$ corresponds to a half-squared filter response at position (x, y) for the opponent channel c; k and σ are the constant scale factor and the semi-saturation constant, respectively. The pool $\sum$ of unit responses considered for the normalization corresponds to units with similar tuning parameters (i.e., orientation, etc) across all color channels c.

All parameters used here are directly constrained by neuroscience data [23,24], which turned out to perform best (i.e., $k = 1$ and $\sigma = 0.225$; see Sec. 4). We found both the rectification and normalization stages to be important for good accuracy (see Table 1).

Double-Opponent (DO) Descriptor: In stage II, DO model unit responses are obtained by filtering SO channels with the spatial sensitivity function $f(x, y)$. Note that, unlike the SO stage, the convolution here is only 2D and $f(x, y)$ contains both center and surround (excitatory/inhibitory) subunits (in the SO computation excitatory and inhibitory subunits are applied on separate color components). With this difference in mind, the spatial sensitivity function $f(x, y)$ used at the DO stage is the same as the one used at the SO stage but in the general case any filter with excitatory and inhibitory components could be used.

Unlike the SO channels, which are normalized by considering pools of units across all color channels, the DO channels are normalized via pools of units at all orientations. Such normalization stage helps sharpen the orientation tuning of the corresponding DO descriptors. The corresponding unit responses are half-squaring rectified and opponent pairs of unit responses further summed to yield invariance to figure-ground reversal. When using oriented filters with multiple phases as in the Hmax model described in Sec. 3, the responses of opponent channels are summed over all phases to yield a phase-invariant DO response (see stage II in Fig. 1).

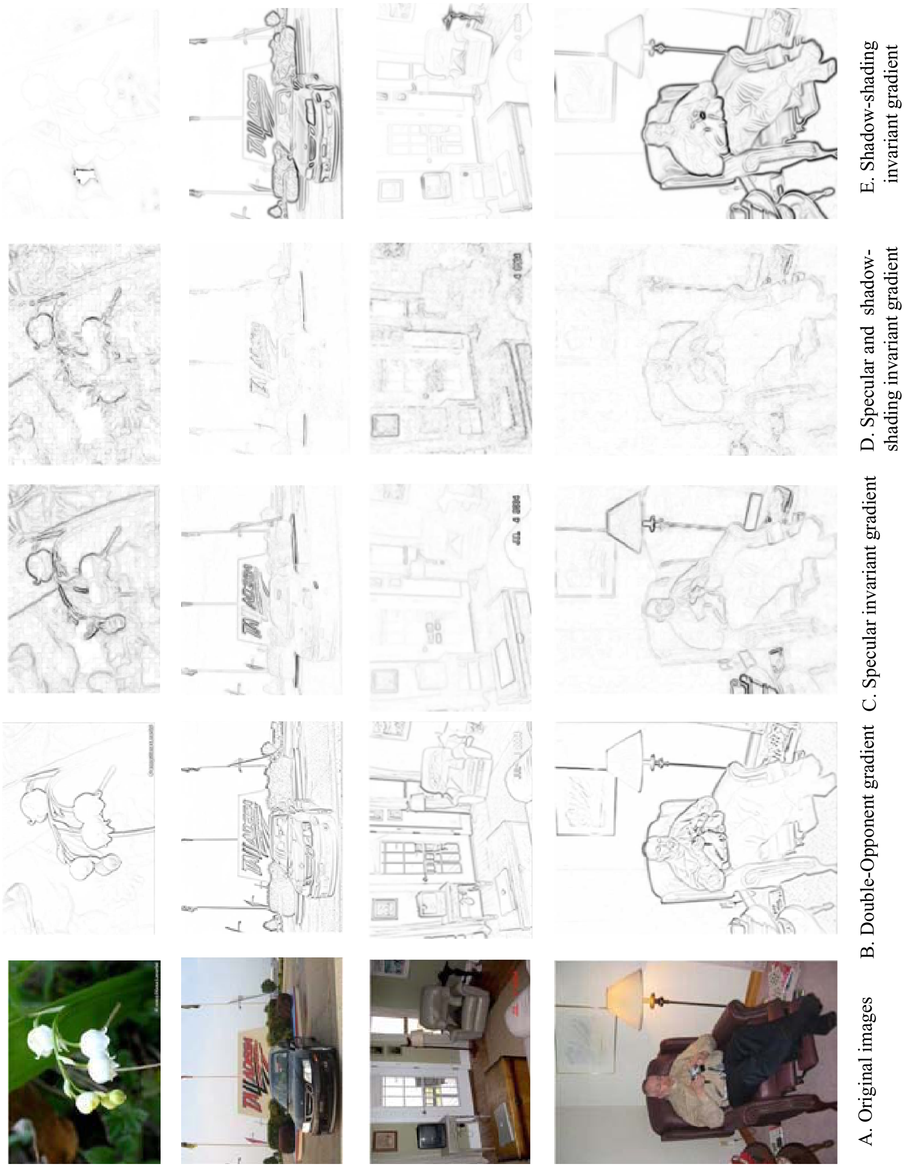

A. Original images
B. Double-Opponent gradient
C. Specular invariant gradient
D. Specular and shadow-shading invariant gradient
E. Shadow-shading invariant gradient

Fig. 3. Comparison between various approaches to color-based gradient computations. (**A**) Original image. (**B**) Proposed Double-Opponent (DO) gradient. (**C**) Specular invariant gradient [25]. (**D**) Specular and shadow-shading invariant gradient [25] (i.e., invariant to both light shift and intensity change). (**E**) Shadow-shading invariant [26] (i.e., invariant to light intensity change). For each method the energy response over multiple orientations is shown.

Fig. 2 shows the responses of the SO and DO stages for a simple image. It is pretty clear that the SO stage mainly contributes to the processing of surface information while the DO stage captures boundary information. Fig. 3 shows representative examples from various state-of-the-art color-based gradient computations (which is a key step in the computation of the color-SIFT descriptors) and a comparison with our proposed DO descriptor.

3 System Extensions

SIFT *Descriptor:* We followed the standard SIFT and bag-of-words implementation described in [20] without the spatial pyramid. The descriptors were computed over a 16 × 16 pixel image patch over a dense grid with a spacing $\varDelta = 8$ pixels. Dense sampling is known to work better than sparse sampling for object and scene recognition [20,27]. K-means was used to cluster the descriptors to form visual words. The size of the codebooks (hard assignment) was determined via cross-validation (leading to 600-2000 centers depending on the dataset). In addition to the grayscale SIFT descriptor, we also implemented two other color descriptors based on the weighted hue (HueSIFT) and opponent angle histograms (OppSIFT) [10] for comparison.

We also considered the OpponentSIFT descriptor that was shown to be the best color descriptor without prior knowledge about the type of light source variation, and the CSIFT descriptor that was the best choice for PASCAL VOC 2007 [6]. In addition, we implemented new SIFT descriptors based on the proposed color processing pipeline: the color tuned SOSIFT and shape tuned DOSIFT as well as their SODOSIFT combination based on the gradients shown in Fig. 1.

HMAX *Model:* We used the standard model implementation [13] with 1,000 features as we have found a moderate improvement with increasing dictionary size. In this extension, the original S1/C1 unit responses at sixteen scales now also included eight SO opponent channels at two orientations[2] and four DO opponent channels at four orientations. An example of the filters used in SOHMAX, DOHMAX and their combination SODOHMAX is shown in Fig. 1.

GIST*:* We extended the GIST algorithm for natural scene categorization [16] from grayscale to color images using the SO and DO descriptors. Filtering in the original algorithm was done in the frequency domain. We thus had to design filters in the spatial domain that would approximate the original system as best as possible. In practice, we found that the Gabor filter parameters used in the HMAX model with six scales (7 × 7 to 39 × 39 in steps of 6 pixels), two phases (0^o and 90^o), and eight orientations ($0 - 180^o$ in steps of 22.5^o) led to results comparable to those obtained with the grayscale GIST model. As an additional benchmark, we considered a color-GIST model computed over RGB images by filtering the R, G, B channels independently.

[2] Because the SO units are only weakly oriented (consistent with electrophysiological studies [23]) we found two orientations to be sufficient.

Berkeley Segmentation Benchmark: We built on earlier work focusing on texture gradient based Gaussian derivatives and center-surround filters [28,17] shown in Fig. 1. We extended the grayscale texture channels with chromatic opponency to obtain color-texture channels at three scales 5, 10, and 20. Fig. 4 shows examples of the resulting texton maps. As suggested in [28], we used local cues and logistic regression to learn the weights for linear cue combinations across scales and hues to obtain the final boundary.

Classification: For all experiments described below, we used LibSVM and the all-pair approach with a χ_2 kernel for the bag-of-words scheme, a linear kernel for the HMAX model, and a Gaussian kernel for the GIST model as done in the original work.

4 Experiments

4.1 Object Recognition

Soccer Team and Flower Classification: Table 2 shows a comparison between the proposed spatio-chromatic (SO and DO) descriptors with other benchmark color and grayscale descriptors on the soccer team dataset (7 classes and 280 images total) [10] and the flower dataset (17 classes and 80 images per class) [14]. Here we followed the standard methodology described in [10,14] for evaluation. Dictionary sizes were optimized for each approach independently using a cross-validation procedure (see Sec. 3). This led to comparable dictionary sizes across descriptor types ($\sim 600 - 1000$ codewords).

On these two datasets, unlike the PASCAL VOC dataset, color cues are highly diagnostic of object category. Individual color descriptors perform better than their grayscale counterparts. The SO and DO descriptors significantly boost the performance of both the SIFT and HMAX (compare the performance under *Color* vs. *Shape* in Table 2). The hue and opponent angle color descriptors (HueSIFT and OppSIFT in Table 2) were previously shown to be the best descriptors for use in combination with a bag-of-words scheme [29].

For a fair comparison, we use the same dense SIFT sampling strategy for all descriptor types. Shown in parenthesis is the performance reported in [10,29] for the HueSIFT and OppSIFT descriptors based on a sparse SIFT sampling strategy (using the Harris-Laplace detector). A late fusion scheme was used here for combining color and shape cues (i.e. SO and DO channels) by concatenating both types of feature responses to form a compact representation. This was found empirically to perform better than an early fusion scheme whereby the two representations were combined into one single descriptor.

We further report the performance of a bag-of-words scheme with a combination of the proposed SO (*Color*) and DO (*Shape*) SIFT descriptors. These two types of descriptors outperform baseline systems both together and in isolation. Interestingly, we found that the HMAX model performs better than the SIFT-based bag-of-words approach (see also [13]).

Table 2. Recognition performance on the soccer team and 17-category flower datasets. Classification accuracy is reported for each feature type (data in parenthesis correspond to the original performance reported in [10,29] using the same features in a bag-of-words scheme.)

	Soccer team			Flower		
Method	Color	Shape	Both	Color	Shape	Both
Hue/SIFT	69 (67)	43 (43)	73 (73)	58 (40)	65 (65)	77 (79)
Opp/SIFT	69 (65)	43 (43)	74 (72)	57 (39)	65 (65)	74 (79)
SOSIFT/DOSIFT	82	66	83	68	69	79
SOHMAX/DOHMAX	87	76	89	77	73	83

We here only compare the performance of bottom-up approaches that do not rely on any prior high-level knowledge related to object categories. It was shown, however, that the performance of various color descriptors could be further improved on this dataset (up to 96% accuracy) when used in conjunction with semantic color features (e.g., color names) and top-down attentional mechanisms [30]. In principle such an approach should similarly improve the performance of the SO and DO descriptors but this should be verified experimentally.

The results obtained on the flower dataset are qualitatively very similar (see Table 2). One small difference is that most shape-based descriptors tend to perform on par or better than their color counterparts. The SO descriptor, which encodes primarily hue information, significantly outperforms the the DO descriptor on the soccer team dataset, which is a predominant color dataset.

Conversely on datasets such as the flower dataset where intra-class variations tend to be large, the DO descriptor which contributes better contours extraction from color cues (as opposed to chrominance information per se) tends to perform best. It has been previously shown that the performance of various descriptors could be further improved with top-down attentional mechanisms (state-of-the-art performance reaching 73% [31] for SIFT descriptors alone and 95% when combined with color names, hue descriptors and SIFT descriptors in the bottom-up and top-down attention framework [30]). Similarly, pre-segmentation and multiple kernel learning methods were shown to further improve performance [32,33,34].

PASCAL VOC *2007 Challenge:* Here we compare the SODOSIFT descriptor (combination of SOSIFT and DOSIFT) on the PASCAL VOC 2007 dataset with other color-based SIFT descriptors as evaluated in [35,6]. Table 3 shows a comparison between the proposed descriptor (i.e. SODOSIFT) and other descriptors using the same bag-of-words implementation as well as published results with the same descriptors (in parenthesis). We also obtain similar performance when incorporating SO and DO into HMAX model. The performance of the SIFT and HMAX with SO or DO extensions are also given on the right (in parenthesis) for comparison.

Table 3. Recognition performance on Pascal voc 2007 dataset. Performance corresponds to the mean average precision (AP) over all 20 classes. Performance (in parenthesis) corresponds to the best performance reported in [35,6].

Method	sift	Huesift	Opponentsift	Csift	SODOsift	SODOHmax
AP	40 (38.4)	41	43 (42.5)	43 (44.0)	46.5 (33.3/39.8)	46.8 (30.1/36.4)

Table 4. Recognition performance on scene categorization

Method	gist	RGBgist	SOgist	DOgist	SODOgist
Accuracy	83.5	84.1	70.5	85.9	87.1

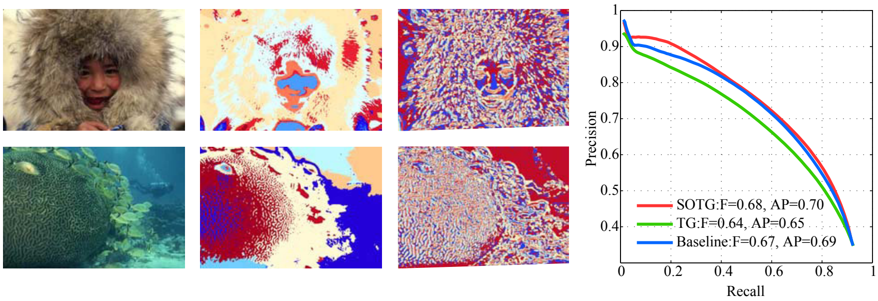
4.2 Scene Categorization

To test our extension of the gist algorithm to color, we used the 8 category scene dataset [16]. Table 4 shows a comparison between the proposed SOgist and DOgist descriptors and their combination SODOgist. We report the average performance over 10 random splits of the data. Unlike the RGBgist and DOgist, which extracts shape information defined by color cues, the SOgist encodes mostly surface properties. The somewhat lower performance of the SOgist on the scene dataset compared to RGBgist and DOgist suggests that color cues may not be diagnostic for the task and that most of the improvement for the RGBgist and DOgist is due to more accurate edge and boundary information.

4.3 Contour Detection

The BSDS500 dataset [17] is an image dataset with human annotations for the evaluation of contour detection and segmentation algorithms. Following the BSDS500 guidelines, precision-recall curves are generated. The best F-measure and the average precision are reported as an overall performance measure for contour detection. Here we built on earlier work that focused on texture gradient using Gaussian derivatives and center-surround filters [28,17].

Here we showed that an extension of the texton-based (grayscale) texture representation by [17] led to a very significant gain in performance (compare TG and its extension SOTG in Fig. 4**B**). The performance of the extended texture channel alone is already higher than the performance of the full system, which combines brightness, color, and texture cues (Baseline in Fig. 4**B**). Fig. 4**A** further shows a comparison between the color-texton maps generated from the texture channel (SOTG; middle column) and the original grayscale texture channel (TG; right column). As can be seen, the proposed color-texture channel seems to capture perceptually more meaningful image regions.



A. Color-texton map vs. grayscale texton map

B. Precision-recall curves

Fig. 4. Contour detection on BSDS500. (**A**) Some representative examples of texton maps and color extensions. From left to right: original images, color-texton map (SOTG) and texton map (TG). (**B**) Precision-recall curves on BSDS500, comparing the original grayscale texture channel with the full Berkeley system that combines brightness, color, and texture cues against our color-texture cue.

5 Conclusion

We have described a hierarchical model of color processing for the extraction of surface (SO descriptor) and boundary/shape (DO descriptor) information based on known properties of the primate visual system and basic canonical circuits (i.e., half-squaring rectification and divisive normalization).

We have further used the framework to extend the HMAX model of object recognition in the visual cortex and a standard bag-of-words model based on the SIFT descriptor. We have tested both approaches on standard image datasets previously used to evaluate color descriptors (soccer team [10] and flower datasets [14]) and object recognition algorithms [15] and shown that the proposed descriptors perform on par or better than other color and shape descriptors. We have further shown increased performance in the context of scene categorization using an extension of the GIST algorithm and contour detection within the Berkeley segmentation system.

Further work is needed to quantitatively assess the invariance properties of the proposed descriptors to changes in illumination. We expect the proposed representation to be tolerant to shifts in light intensity because of the type of filtering used (zero-mean Gabor filters). Furthermore, the normalization used provides some tolerance with respect to small light intensity scaling. The proposed descriptors thus share tolerance properties similar to those of the HueSIFT [25] and the OpponentSIFT descriptors [6]. However the proposed method was shown to yield higher accuracy for object recognition suggesting that the SO and DO descriptors achieve just the right trade-off between selectivity and invariance. Overall the relative success of the proposed biologically inspired approach suggests that neuroscience may contribute new ideas and algorithms for computer vision.

Acknowledgments. This work was supported by grants from DARPA (N10AP20013) and ONR (N000141110743). Additional support was provided by Brown University, the Center for Computation and Visualization, and the Robert J. and Nancy D. Carney Fund for Scientific Innovation. J. Zhang was supported by a grant from the China Scholarship Council (2010669020).

References

1. Hurlber, A.C.: The Computation of Color. Dissertation, Massachusetts Institute of Technology (1989)
2. Wurm, L.H., Legge, G.E., Isenberg, L.M., Luebker, A.: Color improves object recognition in normal and low vision. Journal of Experimental Psychology: Human Perception and Performance 19, 899–911 (1993)
3. Shapley, R., Hawken, M.: Color in the cortex: single- and double-opponent cells. Vision Research 51, 701–717 (2011)
4. Land, E.H., McCann, J.J.: Lightness and retinex theory. Journal of the Optical Society of America 61, 1–11 (1971)
5. Bosch, A., Zisserman, A., Munoz, X.: Scene classification using a hybrid generative/discriminative approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 712–727 (2008)
6. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1582–1596 (2010)
7. Burghouts, G.J., Geusebroek, J.M.: Performance evaluation of local colour invariants. Computer Vision and Image Understanding 113, 48–62 (2009)
8. van de Weijer, J., Gevers, T., Smeulders, A.W.: Robust photometric invariant features from the color tensor. IEEE Transactions on Image Processing 15, 118–127 (2006)
9. Brown, M., Susstrunk, S.: Multi-spectral SIFT for scene category recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 177–184 (2011)
10. van de Weijer, J., Schmid, C.: Coloring Local Feature Extraction. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 334–348. Springer, Heidelberg (2006)
11. Tang, J., Miller, S., Singh, A., Abbeel, P.: A textured object recognition pipeline for color and depth image data. In: International Conference on Robotics and Automation (2012)
12. Gevers, T., Stokman, H.M.G.: Robust histogram construction from color invariants for object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 113–118 (2004)
13. Serre, T., Wolf, L., Bileschi, S.M., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 411–426 (2007)
14. Nilsback, M.E., Zisserman, A.: A visual vocabulary for flower classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1447–1454 (2006)
15. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC 2007) Results (2007), http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html
16. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175 (2001)

17. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 898–916 (2010)
18. Lennie, P., Krauskopf, J., Sclar, G.: Chromatic mechanisms in striate cortex of macaque. The Journal of Neuroscience 10, 649–669 (1990)
19. Conway, B.R.: Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). The Journal of Neuroscience 21, 2768–2783 (2001)
20. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2169–2178 (2006)
21. Heeger, D.J.: Normalization of cell responses in cat striate cortex. Visual Neuroscience 9, 181–197 (1992)
22. Carandini, M., Heeger, D.J., Movshon, J.A.: Linearity and normalization in simple cells of the macaque primary visual cortex. The Journal of Neuroscience 17, 8621–8644 (1997)
23. Johnson, E.N., Hawken, M.J., Shapley, R.: The orientation selectivity of color-responsive neurons in macaque V1. The Journal of Neuroscience 28, 8096–8106 (2008)
24. Solomon, S.G., Lennie, P.: Chromatic gain controls in visual cortical neurons. The Journal of Neuroscience 25, 4779–4792 (2005)
25. van de Weijer, J., Gevers, T., Geusebroek, J.M.: Edge and corner detection by photometric quasi-invariants. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 625–630 (2005)
26. Geusebroek, J.M., van den Boomgaard, R., Smeulders, A.W., Geerts, H.: Color invariance. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 1338–1350 (2001)
27. van Gemert, J., Geusebroek, J.M., Veenman, C.J., Smeulders, A.W.: Kernel Codebooks for Scene Categorization. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 696–709. Springer, Heidelberg (2008)
28. Martin, D., Fowlkes, C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 530–549 (2004)
29. van de Weijer, J., Schmid, C.: Applying color names to image description. In: International Conference on Image Processing, pp. 493–496 (2007)
30. Khan, F.S., van de Weijer, J., Vanrell, M.: Modulating shape features by color attention for object recognition. International Journal of Computer Vision 98, 49–64 (2011)
31. Vigo, D.A.R., Khan, F.S., van de Weijer, J., Gevers, T.: The impact of color on bag-of-words based object recognition. In: International Conference on Pattern Recognition, pp. 1549–1553 (2010)
32. Nilsback, M.E., Zisserman, A.: Automated flower classification over a large number of classes. In: Indian Conference on Computer Vision Graphics Image Processing, pp. 722–729 (2008)
33. Varma, M., Ray, D.: Learning the discriminative power-invariance trade-off. In: IEEE International Conference on Computer Vision, pp. 1–8 (2007)
34. Gehler, P.V., Nowozin, S.: On feature combination for multiclass object classification. In: IEEE International Conference on Computer Vision, pp. 221–228 (2009)
35. van de Sande, K.E., Gevers, T., Snoek, C.G.: Color descriptors for object category recognition. In: European Conference on Color in Graphics, Imaging and Vision, pp. 378–381 (2008)