Biol Cybern
DOI 10.1007/s00422-012-0522-6

LETTER TO THE EDITOR

Biological Cybernetics

# Color coding in the cortex: a modified approach to bottom-up visual attention

**Juan F. Ramirez-Villegas · David F. Ramirez-Moreno**

Received: 16 September 2011 / Accepted: 11 September 2012


**Abstract** Itti and Koch's (*Vision Research* 40:1489–1506, 2000) saliency-based visual attention model is a broadly accepted model that describes how attention processes are deployed in the visual cortex in a pure bottom-up strategy. This work complements their model by modifying the color feature calculation. Evidence suggests that S-cone responses are elicited in the same spatial distribution and have the same sign as responses to M-cone stimuli; these cells are tentatively referred to as red-cyan. For other cells, the S-cone input seems to be aligned with the L-cone input; these cells might be green-magenta cells. To model red-cyan and green-magenta double-opponent cells, we implement a center-surround difference approach of the aforementioned model. The resulting color maps elicited enhanced responses to color salient stimuli when compared to the classic ones at high statistical significance levels. We also show that the modified model improves the prediction of locations attended by human viewers.

**Keywords** Saliency · Visual attention · Double-opponent cell · Center-surround difference · Color map

## 1 Introduction

The influence of color is definitive during visual search (Treisman et al. 1977; Treisman and Gelade 1980; Bergen and Julesz 1983; Desimone and Duncan 1995; Itti and Koch 2001). Humans with normal color vision have three types of retinal cone cells that are tuned to the degree to which long (L), medium (M), and short (S) wavelengths of the spectrum appear in the visual space.

Among mammals, primates have the best color vision (Conway 2009). Color is useful in disambiguating surfaces viewed under ambiguous luminance conditions (i.e., color constancy – assigning colors to objects, irrespective of changes in illumination). Perceptually, color also shows spatial and temporal contrast: red appears redder when surrounded or immediately followed by green (Conway et al. 2002).

To compute color, the brain must compare the relative activation of photoreceptors within approximately the same retinal location. This comparison enables us to disambiguate wavelength and intensity to some extent, but since the resolution of color vision is limited, distinct types of color stimuli generate the same relative cone activation (Gegenfurtner and Kiper 2003; Solomon and Lennie 2007; Conway 2009). Cone signals are processed by various kinds of retinal ganglion cells, most of which project into the lateral geniculate nucleus (LGN).

Saliency maps have been associated with early visual areas with highly structured receptive fields, particularly in the primary visual cortex (area V1). Other approaches have considered overall salience as a consequence of interactions among multiple feature maps, each one codifying the saliency of objects in a specific feature (Itti 2000; Itti and Koch 2000, 2001).

Itti et al. (1998) (and formerly Koch and Ullman 1985) introduced a bottom-up model in which primary visual features are calculated on multiple scales. The color double-opponent system model of their approach has been used several times as a step toward the calculation of dynamic or static saliency (e.g., Bollman et al. 1997; Itti 2004; De Brecht and

Juan F. Ramirez-Villegas and David F. Ramirez-Moreno are contributed equally to the research reported in this work.

J. F. Ramirez-Villegas (✉) · D. F. Ramirez-Moreno
Computational Neuroscience, Department of Physics, Universidad Autónoma de Occidente, Km 2, vía Cali-Jamundi, Cali, Colombia
e-mail: juanfelipe.rv@gmail.com

D. F. Ramirez-Moreno
e-mail: dramirez@uao.edu.co

Published online: 28 September 2012

Saiki 2006; Peters et al. 2005; Gao and Vasconcelos 2007; Walther and Koch 2006; Rapantzikos et al. 2007). In their model, four broadly tuned color channels are created (red, green, blue, and yellow). Four Gaussian pyramids are computed based on these color channels from which the double opponency of cells in V1 is computed by center-surround differences across the color channels. Such a double-opponency calculation aims at reproducing the standard red-green and blue-yellow opponency presumably taking place in the cortex (Engel et al. 1997): while the center of the receptive field of a cell is excited by red (in general, L-cone inputs) and inhibited by green (in general, M-cone inputs), the converse happens in the surround.

Recent findings indicate that S-cone responses are elicited in the same spatial distribution and have the same sign as responses to M-cone stimuli (Conway 2001; Wachtler et al. 2003; Solomon and Lennie 2005); these cells are tentatively referred to as red-cyan (Conway 2001). Likewise, in the case of other cells, the S-cone input seems to be aligned with the L-cone input; these cells might be green-magenta cells (Conway 2001).

This study follows the strategy of calculating color saliency introduced by Itti et al. (1998) – henceforth referred to as the *classic color map* – since the overall mathematical procedure is straightforward, simplified, and approximate, but fast to compute. Our approach calculates green-magenta and red-cyan color double opponencies based on Conway's (2001) findings [and supported by Wachtler et al. (2003) and Solomon and Lennie (2005)]. The proposed color double-opponent system – henceforth referred to as the *modified color map* – leads to the nonlinear enhancement of color saliency and, consequently, to a modified general central representation (modified saliency map) that predicts human eye fixations with better accuracy than the classic approach. Several statistical significance tests were performed to assess the difference between the resulting properties of modified color maps and the classic ones. Furthermore, relevant discussions are provided relating Conway's (2001) experimental evidence to the results of our color models.

## 2 Materials and methods

### 2.1 Modeling approach

#### *2.1.1 Assumptions of the model*

The modified model proposed here follows the assumptions of the classic feature-guided saliency map model by Itti et al. (1998). The main component of this approach relies on the concept of topographical feature maps that represent the overall bottom-up conspicuity of visual stimuli. The attended locations in the visual scene, in general, arise from the computation of such maps. The basic assumptions of this approach have been extensively debated and criticized (see Tatler et al. 2011 for a review). Although Itti et al.'s (1998) approach explains some of the important facts of *attention allocation* (Masciocchi et al. 2009), simple visual features lack the ability to analyze all kinds of information present in highly cluttered real-world scenes (Tatler et al. 2011).

The standard model of bottom-up selective attention by Itti et al. (1998) is limited in many respects. However, the objective of this study is not to present a model in a one-to-one implementation of the brain. Rather, we aimed at (1) presenting a modification of the standard model to improve its plausibility in the view of experimental evidence that has not yet been considered and (2) analyzing how this affects the way in which the model processes visual information.

#### *2.1.2 Primary visual feature extraction (short review)*

As mentioned previously, the feature extraction procedure expands over three main well-studied feature maps: intensity, orientation, and color maps. The center-surround differences are implemented using fine and coarse scales of the Laplacian pyramid (Burt and Adelson 1983) for each feature: the receptive center corresponds to a pixel at a resolution level $c \in \{2, 3, 4\}$, and the surround is the corresponding pixel at the level $s = c + \delta$, with $\delta \in \{3, 4\}$.

#### *2.1.3 Modified color map calculation*

Most LGN cells possess receptive fields similar to those of retinal ganglion cells. Although the role of the LGN in the elaboration of color continues to be controversial, these cells are color selective and, at the same time, are capable of generating opponency (chromatically opponent center and surround signals) but not color contrast calculation. LGN cells mainly come in two flavors: those that compare the relative activation of L and M (loosely called *red-green* cells), and those that compare S activation to some combinations of L and M activations (*blue-yellow* cells). LGN cells are tuned to other cardinal color axes, which are often described as complementary axes (Engel et al. 1997; Schluppeck and Engel 2002; Conway 2009).

LGN cells project into V1 along anatomically segregated streams. Cells in V1 might encode both chromatic and spatial opponency; these cells compare color signals across the visual space. Because of the structure of their specialized receptive fields and the filter matching of color, double-opponent cells are candidates for the neural basis of color contrast (Conway 2009). Some evidence favors two frequent types of double-opponent neurons: L versus M+S (loosely called *red-cyan* cells) and S versus L+M (*blue-yellow* cells). On the other hand, according to Conway (2001), there are also M versus L+S (or *green-magenta*) cells that are perhaps less frequent than *red-cyan* ones.

Double-opponent cells in V1 (Conway 2001; Conway and Livingstone 2006; Wachtler et al. 2003; Johnson et al. 2001; Solomon and Lennie 2005) have receptive fields that resemble the circular, concentric receptive fields of LGN neurons, but both center and surround show color opponency (Gegenfurtner and Kiper 2003; Solomon and Lennie 2007). These cells are chromatically opponent and spatially bandpass for both luminance and isoluminant stimuli (Johnson et al. 2001). A more detailed discussion of these facts can be found in Solomon and Lennie (2007), Gegenfurtner and Kiper (2003), and Schluppeck and Engel (2002).

Conway (2001) established the existence of certain groups of cells sensitive to another type of contrast color. According to his experiments, various types of cells are sensitive to long wavelengths at their center (L) and are simultaneously sensitive to medium and short wavelengths on their surround and vice versa, i.e., for red-on (L+) or yellow-on (S−) stimuli, the center of the cell is excited and at the same time is chromatically opponent to green-on (M+) and blue-on (S+) stimuli (on the surround region). This phenomenon also occurs for green-on center sensitive cells. Both the antagonism and the alignment of M and S cell stimuli are suggestive of the existence of red-cyan and green-magenta cells. Conway's (2001) findings in relation to the coupling of M and S cones in opposition to L-cone inputs in V1 neurons are supported by other studies (Wachtler et al. 2003; Solomon and Lennie 2005).

The response of red-green cells to color stimuli and the fixed ratio of L- and M-cone inputs suggest that these cells encode a single chromatic axis that presumably complements the blue-yellow and the luminance axes. As stated previously, cortical red-green cells usually respond to S-cone stimuli. Hence, the cortical red-green axis might be better described as red-cyan. A red-cyan axis might be advantageous because it and the blue-yellow axis would be silent to shades of gray (Conway 2001). The same phenomenon is observed in the green-magenta axis but, at the same time, is not present in the standard red-green axis [these facts are extended by Conway (2001), who represents the color space as a cube].

Although the cones do not respond to the exposures of red, green, and blue since neural activity peaks are not aligned with these colors in the spectrum, the modeling approach of Itti et al. (1998) functionally approximates their responses by calculating four broadly tuned color channels, namely, $R$, $G$, $B$, and $Y$. Taking into account the aforementioned evidence, we established four different color maps according to these four broadly tuned color channels – $RG(c,s)$, $BY(c,s)$, $RC(c,s)$, and $GM(c,s)$ – to account for red/green, blue/yellow, red/cyan, and green/magenta opponency (respectively), which, in the cortex, are represented by the so-called color double-opponent system:

$$R = \left[\frac{r - (g+b)/2}{I}\right]_+, \tag{1}$$

$$G = \left[\frac{g - (r+b)/2}{I}\right]_+, \tag{2}$$

$$B = \left[\frac{b - (g+r)/2}{I}\right]_+, \tag{3}$$

$$Y = \left[\frac{r + g - 2(|r-g| + b)}{I}\right]_+, \tag{4}$$

where $I$ is the intensity, computed as the mean of the $r$, $g$, and $b$ color bands of the image. Subsequently:

$$RG(c,s) = |(R(c) - G(c))\Theta(G(s) - R(s))|. \tag{5}$$

From Eq. (7) we derive red-cyan double opponent cells as follows:

$$RC(c,s) = |(R(c) + Y(c) - G(c) - B(c)) \Theta(G(s) + B(s) - R(s) - Y(s))|, \tag{6}$$

$$BY(c,s) = |(B(c) - Y(c))\Theta(Y(s) - B(s))|. \tag{7}$$

Finally, green-magenta opponency is calculated as

$$GM(c,s) = |(G(c) + Y(c) - R(c) - B(c)) \Theta(R(s) + B(s) - G(s) - Y(s))|. \tag{8}$$

For all of the preceding equations $R$, $G$, $B$, and $Y$ are the red, green, blue, and yellow broadly tuned color channels, respectively; $R(c)$, $G(c)$, $B(c)$, and $Y(c)$ are the center color signals, and $R(s)$, $G(s)$, $B(s)$, and $Y(s)$ are the surround color signals. The "Θ" symbol describes across-scale subtractions obtained by interpolation to the finer scale and point-by-point subtraction (Itti et al. 1998).

### 2.2 Natural scene data and model evaluation

To measure the performance of modified color maps, i.e., where they stand relative to the well-known model by Itti et al. (1998), static images of natural scenes were gathered from the MSRA Salient Object Database (Liu et al. 2007a,b), image set B, which contains 5,000 images of approximately 300 × 400 pixels.

The images of this database were labeled by a group of nine human observers (all the labeling information is available, along with the database, at Liu et al. 2007b). For each image to be labeled, the database managers asked users to draw a rectangle to enclose the most salient object in the image according to their own understanding. Because the rectangles labeled by different users were not the same, we measured the consistency of the labeling of the images following the procedure detailed by Liu et al. (2007a). Therefore, the saliency probability map is calculated as follows:

$$g_x = \frac{1}{M}\sum_{m=1}^{M} a_x^m, \tag{9}$$

where $M$ is the number of users, $x$ is an index that runs over the entire image set, and $A^m = \{a_x^m\}$ is the binary mask labeled by the $m$th user. The preceding equation represents a saliency probability map since it is the mean of different labeled binary images. Then, to measure the labeling consistency, we computed

$$C_t = \frac{\sum_{x\in\{g_x>t\}} g_x}{\sum_x g_x}, \tag{10}$$

where $C_t$ is the percentage of pixels whose saliency probabilities are above the threshold $t$. Equation (10) describes a proportion intended to provide a measure of consistency of the image labeling across subjects. Values of this parameter close to 1.00 for thresholds close to 1.00 are indicative of high consistency in the labeling process.

Other approaches to measuring the performance of saliency-based models (such as that reported by Masciocchi et al. 2009) include the calculation of correlation measures (e.g., normalized cross-correlation) between the saliency maps and salient locations selected, ordered and marked by a group of subjects according to their own understanding of a set of complex images. Although such an approach is robust in the sense that it directly relates the saliency measures, the approach that we use in this study [first reported by Liu et al. (2007)] is clearly more advantageous, at least in one respect: as the complexity of the scenes increases, the labeling process becomes more inconsistent across subjects, the images produced under such disagreement are discarded, and, consequently, the random sample can be easily limited to contain consistently labeled images. This is not necessarily true for the second approach since there are more variables to be taken into account.

Likewise, under the *location-selection approach* (Masciocchi et al. 2009), interpreting the correlations of the subjects' selected locations to the models' predictions can become obscure and cumbersome: it is expected that the variability across subjects will increase, for instance, in the order of selected locations and, more importantly, the selection of *spurious* interesting locations. In the latter case, determining a relative saliency weight should be a major concern. Additional complications of such an approach may include the selection of suitable criteria to calculate an R-O-C-like measure in order to quantify the performance of the models in predicting human eye fixations.

Consequently, in this study, a random sample of 700 highly consistent labeled images ($C_{0.9} \in \{0.8, 1\}$, i.e., 80.0 % to 100.0 %) was selected for the entire testing procedure in this section. The target object of an image was defined as the set of pixels in the region consistently selected by the subjects. Conversely, the distractors are all pixels outside such a region (note that all regions were selected using $C_{0.9}$). This procedure was carried out with the view toward having an objective performance measurement of the modified color maps for real scenes, given the fact that many studies have classified this operation as a highly subjective process (Itti and Koch 2000; De Brecht and Saiki 2006; Gao and Vasconcelos 2007).

The models were evaluated by computing the standard multiscaled orientation and intensity maps along with the corresponding color maps and then the saliency map. This saliency map as a standard linear combination of the intensity, orientation, and color conspicuity maps (Itti et al. 1998; Itti 2000; Itti and Koch 2001).

Likewise, it should be noted that the color maps are calculated as across-scale sums of the extracted color double-opponent maps, i.e., in the form of standard conspicuity maps.

Let $RG(c,s)$, $BY(c,s)$, $RC(c,s)$, and $GM(c,s)$ be the color maps respectively encoding red-green, blue-yellow, red-cyan, and green-magenta color opponency. The classic color conspicuity map is calculated as

$$\bar{C}_c = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [RG(c,s) + BY(c,s)]\,. \tag{11}$$

Analogously, the modified conspicuity map is computed as

$$\bar{C}_M = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [RC(c,s) + GM(c,s) + BY(c,s)]\,. \tag{12}$$

To this end, using the sample of 700 consistently labeled images we calculated the receiver operating characteristic (ROC) curves of the models with respect to the prediction of eye fixations on natural images (Fig. 5). In general, an ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (1 – specificity) for different testing points (Fawcett 2006). The area under the ROC curve is a measure of the prediction accuracy of a model, i.e., the probability that the algorithm will yield a correct response when a new image is presented. Furthermore, since the prediction rates (accuracy) of either model are proportions, 95% CI can be calculated as $CI(1-\alpha)\% = x \pm Z_{\alpha/2}\sqrt{x(1-x)/n}$, where $\alpha$ is the significance level (typically $\alpha = 0.05$), $Z_{\alpha/2} = 1.96$, $x$ is the mean accuracy, and $n$ is the total number of possible agreements for $x$.

Satisfactory prediction results for classic and modified models were obtained using the two following conditions: (1) The first location attended by the model corresponds to the location of the target (in a pure winner-take-all strategy), and (2) at least 60% of the pixels in the target region present 80% of the maximum saliency value in that region. While the first performance condition is a standard straightforward one (using the so-called winner-take-all, WTA, approach), the second condition aims at measuring objectively the performance of the models based on how homogeneous the activity elicited by salient objects is. Whenever the models did not meet both of these criteria, the result was marked as an unsatisfactory prediction.

## 3 Results

To compare the saliency values between classic color maps (Itti and Koch 2000) and the proposed modified color maps, a Kolmogorov–Smirnov test (KS-test) was performed. Since the KS-test only shows if distributions are different from each other, a one-tailed Student's $t$ test was also carried out to compare the mean values of the distributions. These procedures were performed to compare the ratio of the saliency average values of the targets to the saliency average values of the distractors and to compare the coefficient of variation of activity levels encoded using the classic and modified color map approaches.

The foregoing statistical analyses are graphically supported by empirical cumulative distribution function (CDF) plots, box plots, and comparison plots for the mean distribution values (Figs. 1 and 2). In particular, the box plots were drawn as follows. The top and bottom sections of each box are the 25th and 75th percentiles of the samples, respectively. The line in the middle of each box is the sample median. The dashed lines extending below and above each box are drawn from the ends of the interquartile ranges to the furthest observation within the dashed line length. The crosses represent the outliers in the samples, i.e., atypical data sufficiently distant from the limits of the box.

The results of the two statistical procedures are summarized in Fig. 1. The values of the abscissa in Fig. 1a are the target/distractor saliency ratios. Analogously, the values of the abscissa, $x$, in Fig. 1b are the values of the coefficient of variation of activity levels elicited by targets. In both cases, the values of the ordinate, $F(x)$, are the cumulative frequencies at which each value of the abscissa is observed. Each plot depicts two distributions, one corresponding to the classic, and the other corresponding to the modified color map. The distributions of the saliency levels of the targets in the classic and modified color maps showed significant differences ($p < 10^{-5}$); moreover, as can be noticed from the location of the empirical cumulative distribution plots and the comparison plots, the classic color maps led to higher target/distractor saliency ratio levels than modified ones (with mean values of $2.117 \mp 0.0728$ and $1.763 \mp 0.0728$, respectively; see Fig. 1a, bottom, and Fig. 2 for the semi-log plot of the cumulative distribution). One important fact of these cumulative distributions is that, despite their locations, the shape of the distributions is very similar, even though they are not linearly dependent on each other.

Furthermore, from Fig. 1 two main facts can be noticed. First, the classic maps encode a higher ratio between target saliency and distractor saliency than the modified maps, which indicates that discrimination in the former map is better. Second, the modified maps encode fully enhanced color axes for targets and distractors, making the activity of objects more homogeneous. The last fact is clearly shown by the coefficient of variation distribution values (Fig. 1a): as the standard deviation $s \to 0$, the activity elicited by a given object tends to be completely homogeneous. Furthermore, the small gap between the CDFs in Fig. 1a (ranging approximately from 1.00 to 6.00) produces a less significant effect than the effect observed in the CDFs for the coefficient of variation. In addition, from the box plots depicted in Fig. 1b, it can be noticed that modified maps tend to produce less variability across the image sample than the classic ones, which indicates that they might be more invariant to different properties inherent in natural images. This all improves the overall accuracy of the model in terms of the prediction of human eye fixations.

The difference between the responses elicited by the models can be noticed from the CDFs. The KS-test and the Student's $t$ test were also carried out to measure the differences in the coefficient of variation (the ratio of the standard deviation to the mean) of activity levels encoded using the classic and modified color map approaches. We used this parameter to measure the uniformity/homogeneity of the activity levels of the targets distributed in the two types of color maps; in this way, two objects could coincidentally have the same activity standard deviation but, at the same time, different mean levels of activity. Therefore, it is useful to have a normalized measure of the distributions, in other words, the standard deviation of the activity levels must be explained in the context of the mean. The results show that there are significant statistical differences ($p < 10^{-50}$) between the aforementioned color maps (for both the distribution and the mean value). Moreover, as expected, the coefficient of variation values for classic maps were greater than those for modified maps (with mean values of $0.5440 \mp 0.0079$ and $0.4190 \mp 0.0079$, respectively; see Fig. 1, bottom). This illustrates that the saliency of objects is more dispersed in classic maps (Fig. 3).

In light of these results, there are at least two important properties that emerge from the modified color maps in overall saliency computing. On the one hand, modified cell maps (those accounting for red-cyan and green-magenta axes) elicit stronger responses for both targets and distractors, and this is an expected effect of our model (color contrast properties are enhanced for the whole visual content of the natural scenes and the color contrast is calculated using a more complete set of axes for color contrast, i.e., another fully enhanced output nonlinearity). On the other hand, modified cell maps show more homogeneous responses to targets and distractors, and at this point our color map model strongly outperforms the outcomes of the standard model by Itti et al. (1998).

Some of the main results regarding the visual differences between color maps are depicted in Figs. 3 and 4. As can be seen, modified color maps in this case induce both higher and more uniform color saliency for targets, thereby improving the overall performance of the model in detecting salient objects in natural scenes.

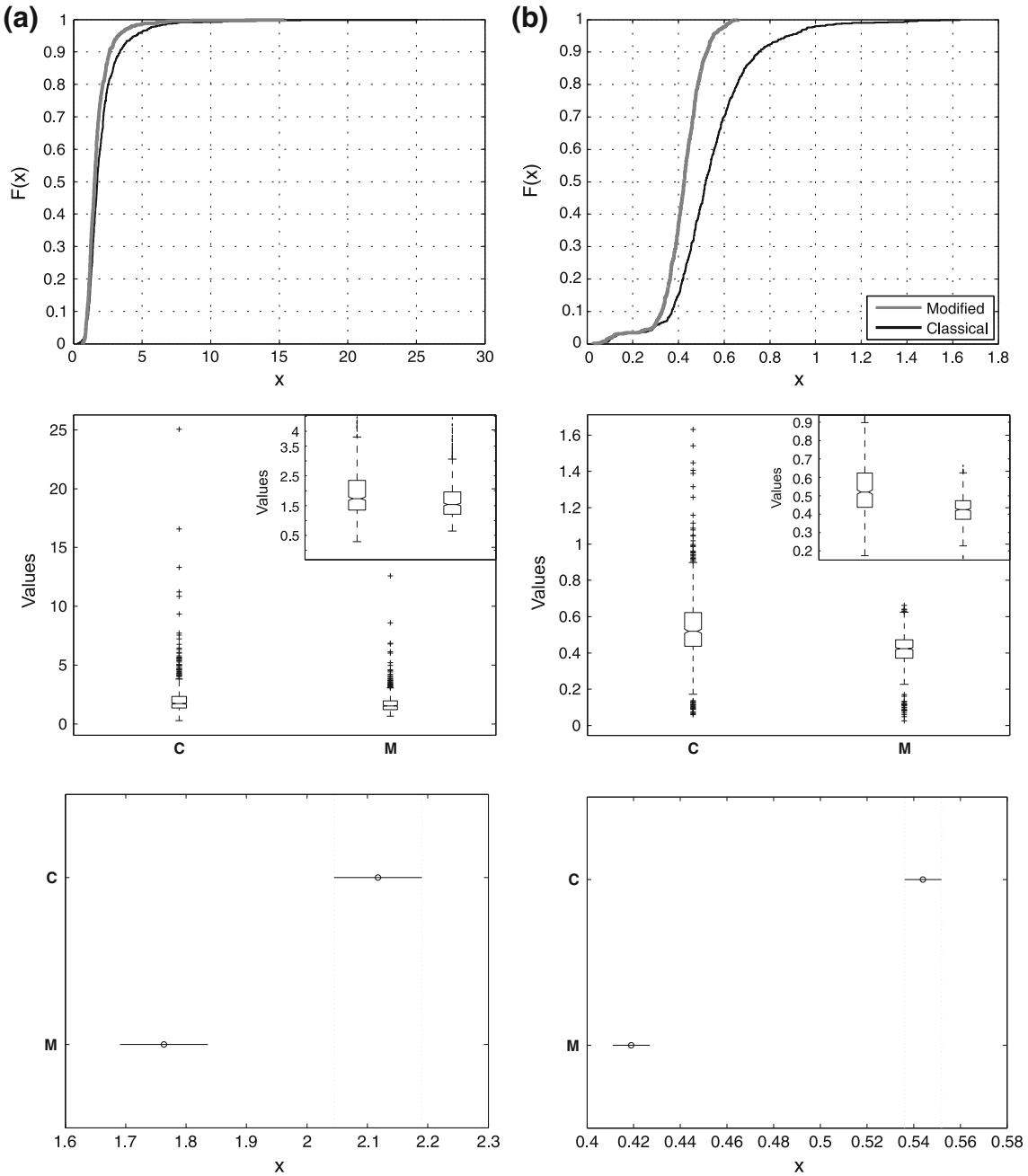


**Fig. 1** Statistical results of modified (M) and classic (C) color maps on saliency map calculation of the random sample of 700 highly consistent labeled images: **a** cumulative distribution function (*top*), box plots (zoomed in at upper right corner) (*middle*), and comparison plot (*bottom*) of saliency ratio target/distractors; **b** cumulative distribution function (*top*), box plots (zoomed in at upper right corner) (*middle*), and comparison plot (*bottom*) of coefficient of variation of activity levels of targets. For all purposes, $F(x)$ is the cumulative distribution of the values of the random sample and $x$ runs over the distribution values. The multiple comparison test plots in **a** and **b** (*bottom*) illustrate the mean of each set/group (corresponding to either the modified or classic color map measurements) and the standard error bars around the mean. In both cases, the means were significantly different since their intervals did not overlap ($p < 10^{-5}$ for the saliency ratio and $p < 10^{-50}$ for the coefficient of variation). Furthermore, the 95% confidence intervals (CIs) for the difference between the means were (0.2089–0.4999) and (0.1092–0.1408) for **a** and **b**, respectively

In this way, under the particular setup described in Sect. 2.2 (60% of the pixels at 80% of the maximum saliency value), the performance (with a 95% CI) of the visual-saliency-based model increased from 61.97% (58.00–65.21) to 74.14% (70.82–77.32) (accuracy) in comparing Itti et al.'s (1998) standard model and our model.

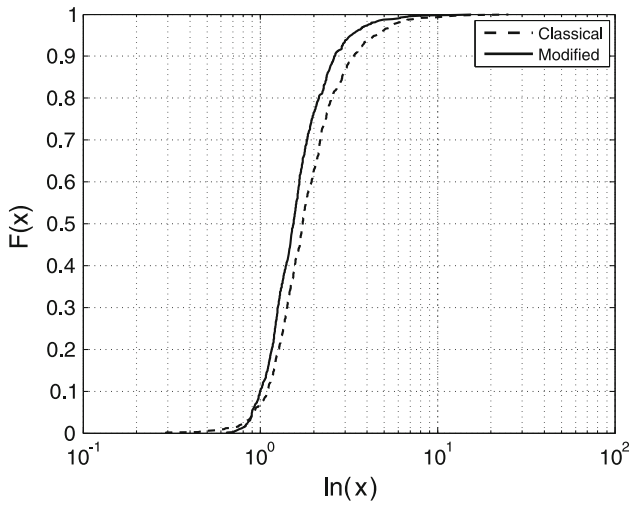


**Fig. 2** Cumulative distribution function of logarithmic saliency ratio target/distractors. For all purposes, $F(x)$ is the cumulative distribution of the values of the random sample, and $x$ runs over the distribution values

Additional experiments indicated that the comparative performance of the models was dependent on the thresholds chosen (not shown in Fig. 5). In general, it is expected that the sensitivity will increase as both percentages decrease. The results of these experiments can be summarized as follows:

(1) Decreasing either of the percentage values, in general, raised the performance of the classic approach (by a maximum of ∼7% when the second performance criterion was completely disregarded), while the performance of the modified approach was raised moderately (by a maximum of ∼2% when the second performance criterion was completely disregarded).
(2) Increasing either of the percentage values produced a significant drop in the performance of the classic approach (by a maximum of ∼11% for 90% of the pixels at 90% of the maximum saliency value) and, at the same time, such an increase moderately decreased the performance of the modified approach (by a maximum of ∼ 6% for 90% of the pixels at 90% of the maximum saliency value).

Additionally, it should be noted that the large differences in the curves obtained from ROC analysis arise from different intrinsic features of the color map models (since they are essentially two very different nonlinear outputs) and from the particular way in which we measured the overall

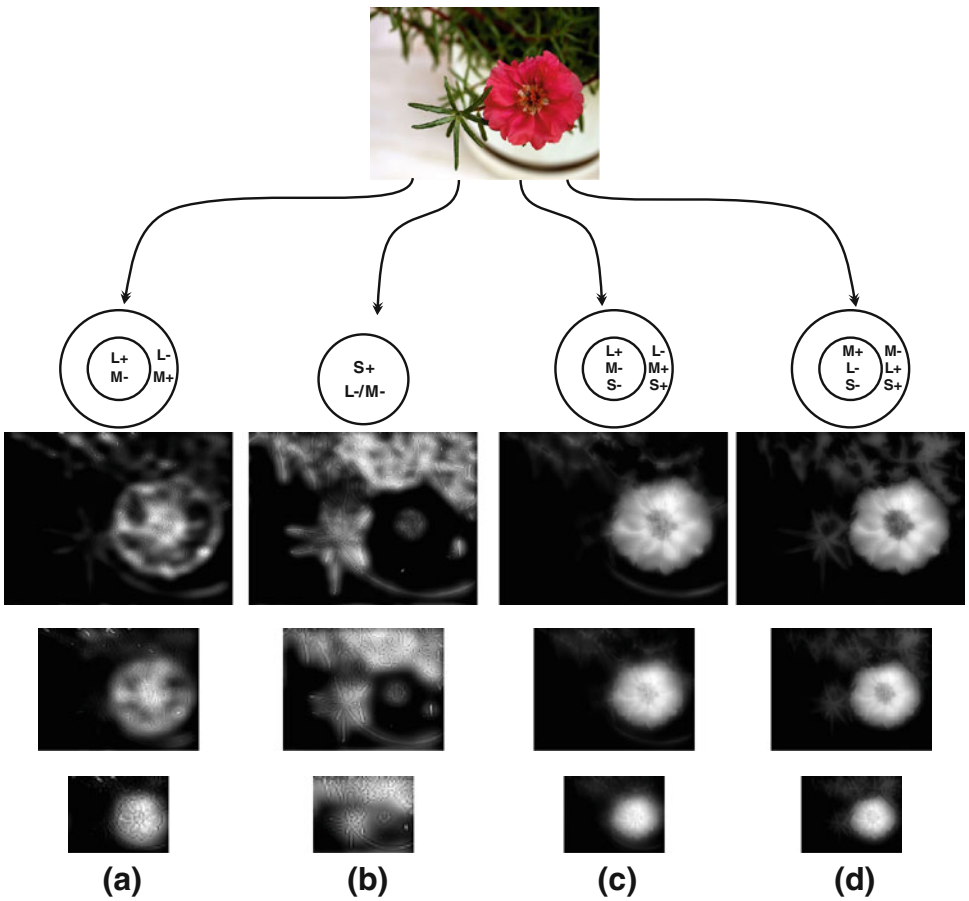


**Fig. 3** Color feature maps: **a** standard red-green cell model; **b** blue-yellow cell model; **c** modified red-cyan cell model; **d** modified green-magenta cell model. (Color figure online)

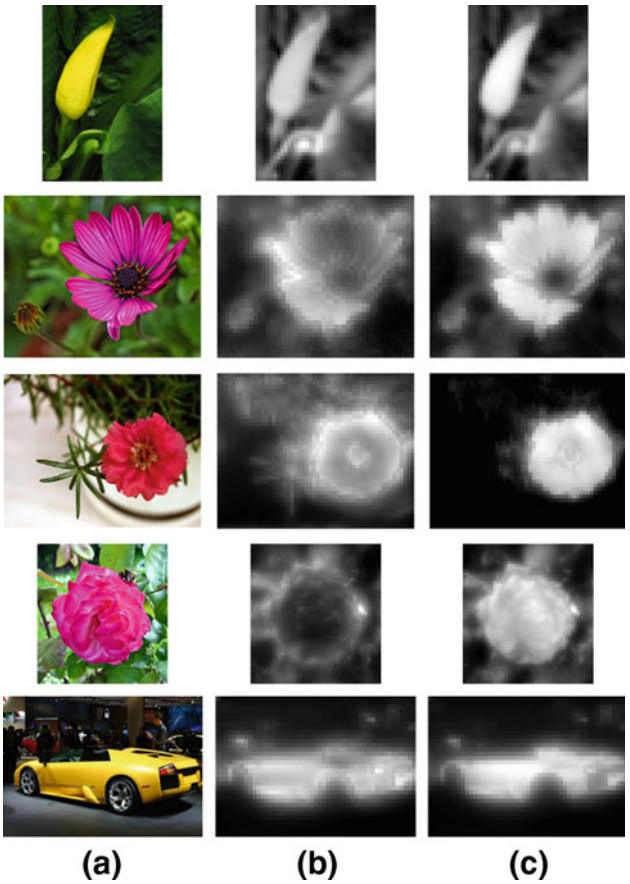


**Fig. 4** Color maps calculated as across-scale sums of corresponding color double-opponent channels: **a** input images; **b** classic color maps (*RG* + *BY*); **c** modified color maps (*RC* + *BY* + *GM*). (Color figure online)

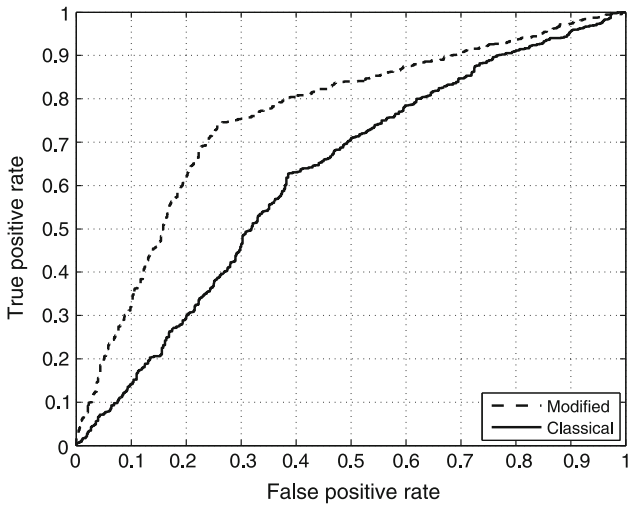


**Fig. 5** ROC curve results of modified ($A = 0.7414$) and classic ($A = 0.6197$) color maps on saliency map calculation for predicting human eye fixations using a random sample of 700 consistently labeled images

performance of the models. The fact that the distributions of target-to-distractor saliency ratios of both models are closely related (Fig. 1a) does not necessarily mean a similarity per se on the output either for target selection or for the homogeneity of the saliency levels of the objects at any level of the color map representations.

The results obtained from ROC analysis are consistent with the large difference between the distributions found for the coefficient of variation of activity levels encoded by classic and modified color map approaches (Fig. 1).

## 4 Discussion and final remarks

Like other features, color processing involves a series of hierarchical steps that begin with three cone classes. Color signals go from retinal ganglion cells through LGN and V1. Highly specialized double-opponent cells in V1 compute color contrast and color constancy along different color axes. According to the evidence established by Conway (2001) and supported by other relevant studies (e.g., Wachtler et al. 2003; Solomon and Lennie 2005), these color axes can be loosely called red-cyan, blue-yellow, and green-magenta.

As mentioned previously, color representation was modified in this work by incorporating green-magenta and red-cyan cells regarding color-opponency mechanisms, based on experimental data obtained by Conway (2001). According to our findings, in this particular regard, the Itti and Koch (2000) color model needed revision, as color feature codification often failed to distinguish salient color objects when in analyses of complex scenes; this was sufficiently illustrated by our results. Other experimental evidence supports the existence of color cells often referred to as *modified type II* cells (T'so and Gilbert 1998). These have been defined as having a color-opponent center and a broadband (nonchromatically) surround that suppresses any effect on the center. According to Conway (2001, 2009), however, there is no evidence that clearly demonstrates this fact as most color cells in V1 are actually double opponent and respond with different intensities to color stimuli in the surrounds. Given these facts, we did not take into account these modified type II cells in our model.

Our model does not account for the fact that the three types of color-sensitive photoreceptors in the human retina have their peak sensitivity at light wavelengths that are not matched to the primary colors (Itti et al. 1998). In neurobiological terms, this implies that the cones do not correspond to the sensations of red, green, and blue, whereas in reality, even different cells of the same cone class can produce different color sensations (Hofer et al. 2005). Therefore, the representations of color signals in this work are likely to be gross approximations of real color signals.

In addition, various approaches (including ours) have focused on basic color features in order to address how visual attention is directed toward a bottom-up or a feature-guided strategy. However, as far as the computational approaches indicate, modeling efforts have been concentrated on the LGN (wired up to detect luminance contrast, not color contrast) and V1 (often referred to as the neural basis for color contrast and color constancy) (Conway 2009; Solomon and Lennie 2005; Gegenfurtner and Kiper 2003). Other experimental evidence focuses on the role of V2, the posterior inferior temporal cortex (PIT), and the inferior temporal cortex (IT) in the elaboration of color (Conway 2009). To this end, several issues regarding color codifying and its interpretation continue to be computationally challenging problems and – to our knowledge – remain to be addressed. These issues should be matters of future research and modeling efforts.

## References


Bergen JR, Julesz B (1983) Parallel versus serial processing in rapid pattern discrimination. Nature 303:696–698

Bollman M, Hoischen R, Mertsching B (1997) In: Berlin et al (eds) Integration of static and dynamic scene features guiding visual attention. Springer, pp 483–490

Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Trans Commun 31:532–540

Conway BR (2001) Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). J Neurosci 21:2768–2783

Conway BR (2009) Color vision, cones and color-coding in the cortex. Neuroscientist 15:274–290

Conway BR, Livingstone MS (2006) Spatial and temporal properties of cone signals in alert macaque primary visual cortex. J Neurosci 26:10826–10846

Conway BR, Hubel DH, Livingstone MS (2002) Color contrast in macaque V1. Cereb Cortex 12:915–925

De Brecht M, Saiki J (2006) A neural network implementation of a saliency map model. Neural Netw 19:1467–1474

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222

Engel S, Zhang X, Wandell B (1997) Colour tuning in human visual cortex measured with functional magnetic resonance imaging. Nature 388:68–71

Fawcett T (2006) An introduction to ROC analysis. Pat Rec Lett 27:861–874

Gao D, Vasconcelos N (2007) Bottom-up saliency is a discriminant process. Proceedings of the IEEE international conference on computer vision

Gegenfurtner KR, Kiper DC (2003) Color vision. Annu Rev Neurosci 26:181–206

Hofer H, Singer B, Williams DR (2005) Different sensations from cones with the same photopigment. J Vis 5:444–454

Itti L (2000) Models of bottom-up and top-down visual attention. Dissertation, California Institute of Technology, Pasadena, CA

Itti L (2004) Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Trans Image Process 13:1304–1318

Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res 40:1489–1506

Itti L, Koch C (2001) Computational modeling of visual attention. Nat Rev Neurosci 2:194–203

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intel 20:1254–1259

Johnson EN, Hawken MJ, Shapley R (2001) The spatial transformation of color in the primary visual cortex of the macaque monkey. Nat Neurosci 4:409–416

Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219–227

Liu T, Sun J, Zheng NN, Tang X, Shum HY (2007a) Learning to detect a salient object. Proceedings of IEEE computer society conference on computer and vision pattern recognition

Liu T, Sun J, Zheng NN, Tang X, Shum HY (2007b) MSRA Salient Object Database. Microsoft Research. http://research.microsoft.com/en-us/um/people/jiansun/salientobject/salient_object.htm. Accessed 12 June 2012

Masciocchi CM, Mihalas S, Parkhurst D, Niebur E (2009) Everyone knows what is interesting: salient locations which should be fixated. J Vis 9:1–22

Peters RJ, Iyer A, Itti L, Koch C (2005) Components of bottom-up gaze allocation in natural images. Vis Res 45:2397–2416

Rapantzikos K, Tsapatsoulis N, Avrithis Y, Kollias S (2007) Bottom-up spatiotemporal visual attention model for video analysis. Image Process IET 1:237–248

Solomon SG, Lennie P (2005) Chromatic gain controls in visual cortical neurons. J Neurosci 25:4779–4792

Solomon SG, Lennie P (2007) The machinery of colour vision. Nat Rev Neurosci 8:276–286

Schluppeck D, Engel SA (2002) Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording. J Vis 2:480–492

T'so DY, Gilbert CD (1988) The organization of chromatic and spatial interactions in the primate striate cortex. J Neurosci 8:1712–1727

Tatler BW, Hayhoe MM, Land MF, Ballard DH (2011) Eye guidance in natural vision: reinterpreting salience. J Vis 11:1–23

Treisman A, Sykes M, Gelade G (1977) Selective attention stimulus integration. Lawrence Erlbaum Associates, Hillsdale, pp 333–361

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognit Psychol 12:97–136

Wachtler T, Sejnowski TJ, Albright TD (2003) Representation of color stimuli in awake macaque primary visual cortex. Neuron 37:681–691

Walther D, Koch C (2006) Modeling attention to salient proto-objects. Neural Netw 19:1395–1407