Saliency of color image derivatives: a comparison between computational models and human perception

Eduard Vazquez,[1,*] Theo Gevers,[2] Marcel Lucassen,[2] Joost van de Weijer,[1] and Ramon Baldrich[1]

[1]*Computer Vision Center, Universitat Autónoma de Barcelona, 08193 Cerdanyola del Valles, Barcelona, Spain*
[2]*Faculty of Science, University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands*
[*]*Corresponding author: eduard.vazquez@cvc.uab.cat*

Received June 19, 2009; revised December 22, 2009; accepted December 25, 2009;
posted January 19, 2010 (Doc. ID 112943); published February 26, 2010

In this paper, computational methods are proposed to compute color edge saliency based on the information content of color edges. The computational methods are evaluated on bottom-up saliency in a psychophysical experiment, and on a more complex task of salient object detection in real-world images. The psychophysical experiment demonstrates the relevance of using information theory as a saliency processing model and that the proposed methods are significantly better in predicting color saliency (with a human-method correspondence up to 74.75% and an observer agreement of 86.8%) than state-of-the-art models. Furthermore, results from salient object detection confirm that an early fusion of color and contrast provide accurate performance to compute visual saliency with a hit rate up to 95.2%. © 2010 Optical Society of America

OCIS codes: 330.1720, 110.2960, 330.5510, 150.1135, 330.1880.

1. INTRODUCTION

Human visual attention is for an important part bottom-up driven by the saliency of image details. An image detail appears salient when one or more of its low-level features (e.g., size, shape, luminance, color, texture, binocular disparity, or motion) differs significantly from its variation in the background. Saliency determines the capability of an image detail to attract visual attention (and thus guide eye movements) in a bottom-up way [1,2]. Current models of human visual search and detection suggest that this preattentive stage indicates potentially interesting image details, whereupon the focus of attention is sequentially shifted to each of these regions and the serial stage is deployed to analyze them in detail [3]. Computational saliency could correspondingly assist in the efficient assessment of image content.

Computational saliency models based on information theory have been shown to successfully model human saliency from local image features [4–6]. This theory states that feature saliency is inversely related to feature occurrence, i.e., rare features are more informative and therefore more salient than features that occur more frequently. It is indeed plausible that interesting image details correspond to locations of maximal information content, a measure closely related to local feature contrast [7,8]. Consequently, recent models of human visual fixation behavior assume that saliency-driven free viewing corresponds to maximizing information sampling [9,10]. These models have successfully been deployed to model human fixation behavior, pop-out, dynamic saliency, saliency asymmetries, and to solve classic computer vision problems like dynamic background subtraction [9,8,11].

Because of its importance for many practical applications, we focus on bottom-up saliency in this paper. The parallel, preattentive, or bottom-up stage of human vision is thought to guide a serial (computationally intensive) attentive or top-down stage. Among all features that contribute to a detail's saliency, orientation and color are generally considered to be the most significant ones [12–14]. Consequently, most current saliency models are based on local color and orientation contrast (e.g., [15–18]). In general, individual saliency maps for these features are computed. Subsequently these maps are merged in a late stage into a single overall saliency map, also called late fusion of features [16]. However, there exist evidence that the human visual system combines low-level features in an early stage [2,19]. Information theoretical methodology can be used to compute the saliency of color edges by combining chromaticity and contrast in an early stage.

Therefore, in this paper, a method is proposed that computes image saliency from the information content (the frequency of occurrence) of chromatic derivatives. The method is based on the observation that in natural images, color transitions of equal probability (i.e., isosalient transitions) form ellipsoids in decorrelated color spaces [20]. The transformation that turns these ellipsoidal isosaliency surfaces into spherical ones (called the color saliency function) ensures that vectors of equal length have equal information content and thus equal impact on the saliency function. In [20] the statistics of the color transitions are based on a collection of images. In addition, we investigate transformations based on color transition statistics of a single image.

To investigate the correspondence between our computational saliency model and human visual perception, we perform a psychophysical experiment. The aim of the experiment is to verify the proposed model on a purely

bottom-up saliency task in a controlled environment where possible effects of cognitive top-down mechanisms are avoided. As a second experiment we validate our saliency model on the task of salient object detection in real-world images and compare it with existing methods. We call this task *high-level saliency* to distinguish it from saliency detection that is purely bottom-up. In high-level saliency tasks, the bottom-up saliency is complemented with top-down mechanisms, such as image semantics. As bottom-up color saliency is one of the mechanisms contributing to high-level saliency, a correlation between them is to be expected.

The paper is organized as follows. In Section 2, three computational saliency methods are proposed, complemented by a multiscale approach. In Section 3, the bottom-up saliency is evaluated by a psychophysical experiment. In Section 4, the proposed computational methods are evaluated on the high-level saliency task of salient object detection in real-world images. Finally, in Section 5 conclusions are drawn.

2. SALIENCY OF COLOR EDGES

In this section, three different computational methods are presented to compute color edge saliency based on the information content of color edges: (1) a local version that estimates color edge saliency from a single image, (2) a global version that uses a collection of images to compute color edge saliency, and (3) a version in which the eigenvectors of the transformation matrix are restricted to the opponent color space [see Eq. (7)]. Finally, a multiscale approach is presented to improve saliency detection in real-world images.

A. Computational Saliency Measure Based on Chromatic Transitions

The color saliency method by Van de Weijer *et al.* [20] is inspired by the notion that a feature's saliency reflects its information content as follows. Consider an image $\mathbf{f}=(R,G,B)^{\mathbf{t}}$. The information content, I, of an image derivative $\mathbf{f}_{\mathbf{x}}$, according to *information theory*, is given by the logarithm of its probability p:

$$I=-\log[p(\mathbf{f}_{\mathbf{x}})]. \tag{1}$$

Hence, color image derivatives that are equally frequent, from now on named isosalient derivatives, have equal information content. In Fig. 1, the distribution of RGB derivatives for the 40,000 images of the COREL data set is given. The isosalient derivatives form an ellipsoidlike distribution, of which the longest axis is along the luminance direction. This indicates that equal displacements (i.e., points with equal norm of the chromatic derivatives) are more informative along the color directions (perpendicular to the luminance) than in the luminance direction.

To map image derivatives to a saliency map, a function g is required that maps isosalient derivatives to equal saliency. We choose to map the derivatives to a new space where isosalient derivatives have equal norms:

$$p(\mathbf{f}_{\mathbf{x}})=p(\mathbf{f}'_{\mathbf{x}})\leftrightarrow|g(\mathbf{f}_{\mathbf{x}})|=|g(\mathbf{f}'_{\mathbf{x}})|,$$

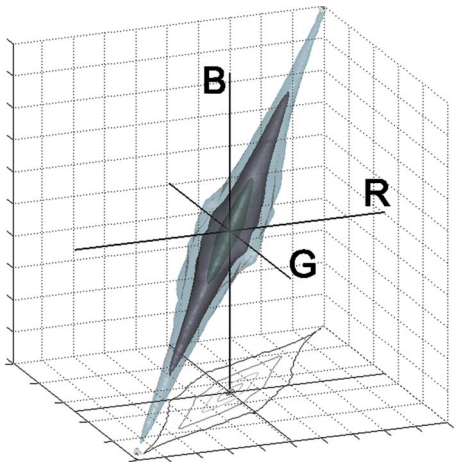

Fig. 1. (Color online) Histogram of the distribution of RGB derivatives computed for the 40,000 images of the COREL image data set. The iso-salient derivatives form an ellipsoidlike distribution, of which the longest axis is along the luminance direction.

$$p(\mathbf{f}_{\mathbf{x}})<p(\mathbf{f}'_{\mathbf{x}})\leftrightarrow|g(\mathbf{f}_{\mathbf{x}})|>|g(\mathbf{f}'_{\mathbf{x}})|. \tag{2}$$

The function g is called the saliency transformation, whose norm can now be interpreted as the saliency. Note that this only puts a weak constraint on the color saliency functions g. It is required that derivatives with equal information content are mapped to vectors with equal norm. A more restrictive constraint would be to require the saliency function to map derivatives to a space in which their norm is proportional to their information content. This further improves saliency detection. However, this is not a uniquely color phenomenon and holds as well for luminance contrast saliency.

Here we model the surface of isosalient derivatives with an ellipsoid. We estimate the parameters of the ellipsoid by the covariance matrix $\mathbf{N}$:

$$\mathbf{N}=\overline{\mathbf{f}_{\mathbf{x}}(\mathbf{f}_{\mathbf{x}})^{\mathbf{t}}}=\begin{pmatrix}\overline{R_{\mathbf{x}}R_{\mathbf{x}}} & \overline{R_{\mathbf{x}}G_{\mathbf{x}}} & \overline{R_{\mathbf{x}}B_{\mathbf{x}}}\\ \overline{R_{\mathbf{x}}G_{\mathbf{x}}} & \overline{G_{\mathbf{x}}G_{\mathbf{x}}} & \overline{G_{\mathbf{x}}B_{\mathbf{x}}}\\ \overline{R_{\mathbf{x}}B_{\mathbf{x}}} & \overline{G_{\mathbf{x}}B_{\mathbf{x}}} & \overline{B_{\mathbf{x}}B_{\mathbf{x}}}\end{pmatrix}, \tag{3}$$

where the matrix elements are computed by

$$\overline{R_{\mathbf{x}}R_{\mathbf{x}}}=\sum_{i\in S}\sum_{\mathbf{x}\in X^i}R_{\mathbf{x}}(\mathbf{x})R_{\mathbf{x}}(\mathbf{x}), \tag{4}$$

where S is a set of images, and X^i is the set of pixel coordinates $\mathbf{x}$ in image i. Matrix $\mathbf{N}$ describes the derivative energy in any direction $\hat{v}$. This energy is computed by $E(\hat{v})=\hat{v}^{\mathbf{t}}\mathbf{N}\hat{v}$. Matrix $\mathbf{N}$ can be decomposed into eigenvector matrix $\mathbf{U}$ and eigenvalue matrix $\mathbf{\Lambda}$ according to $\mathbf{N}=\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}$. This provides the saliency function g:

$$\mathbf{g}(\mathbf{f}_{\mathbf{x}})=\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{f}_{\mathbf{x}}. \tag{5}$$

Substitution of Eq. (5) into Eq. (3) yields

$$\mathbf{g}(\mathbf{f}_{\mathbf{x}})[\mathbf{g}(\mathbf{f}_{\mathbf{x}})]^{\mathbf{t}}=\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}\mathbf{U}\mathbf{\Lambda}^{-1}=I, \tag{6}$$

meaning that the covariance matrix of the transformed image is equal to the identity matrix. This implies that the derivative energy in the transformed space is equal in all directions.

In this paper, three computational saliency methods are considered that are derived from information theory. The first method corresponds to the one introduced in [20] and will be evaluated in Section 3 with a psychophysical experiment. Additionally, the aim is to investigate the performance of two new methods to compute the color saliency transformation.

- *Global opponent color-space saliency* [20]: Saliency is defined as the rarity of color derivatives in a set of images, with the additional restriction that the eigenvectors of the saliency matrix coincide with the vectors that span the opponent color space. In this case,

$$\mathbf{U}^{\mathbf{t}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}. \tag{7}$$

The opponent color space decorrelates the chromatic channels from the luminance channel. The first channel is the red–green channel, the second the blue–yellow channel, and the third the luminance component. The color saliency transformation $\mathbf{M}_o^c = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}$ only differs in the scaling of the axes as given by the eigenvalue matrix $\mathbf{\Lambda}$, whose values can be computed with $\mathrm{diag}(\mathbf{\Lambda}) = \mathrm{diag}\{\mathrm{sqrt}[\overline{\mathbf{U}^{\mathbf{t}}\mathbf{f}_{\mathbf{x}}(\mathbf{U}^{\mathbf{t}}\mathbf{f}_{\mathbf{x}})^{\mathbf{t}}}]\} = (\alpha, \beta, \gamma)$, computed similarly as in Eqs. (3) and (4). The diag function reduces a matrix to its diagonal elements. The same eigenvalue matrix is applied to all images.

- *Global color saliency:* Here, saliency is defined as the rarity of the color derivatives over a set of images. Hence, a single matrix $\mathbf{M}_g^c$ is computed based on the color derivatives of all images in a data set (S contains all images). The same saliency matrix is then applied to all images in the data set.
- *Local color saliency:* Saliency is defined by the rarity of the color derivatives in a single image. Thus, when applied to a set of images, each image is transformed by its own individual saliency matrix $\mathbf{M}_l^c$ (where c stands for computational and l for local). For its computation, S in Eq. (4) contains only a single image.

An example of local and global computational saliency is given in Fig. 2. Based on global saliency, the edges of the red American flag are considered salient. However, for local saliency, which is computed from the statistics of this image, the red edges are not considered salient. Instead the brown edges of the pastry are considered more salient. This is in correspondence with human assessment of this image [21].

B. Multiscale Color Saliency

In natural scenes, salient regions appear at multiple scales. For this reason we propose to extend our computational saliency models with a multiscale approach. Maps computed at multiple scales can be combined into a single saliency map as follows:

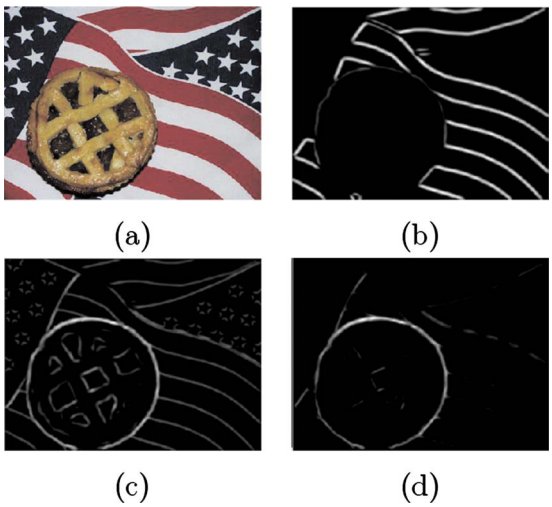

Fig. 2. (Color online) (a) Original image, (b) global color saliency, (c) local color saliency, (d) multiscale local color saliency. Global saliency amplifies the red edges of the flag. Based on the local statistics of this image, the local saliency increases the saliency of the pie. The multiscale approach suppresses the colorful edges of the American flag further; therefore, the pie is better detected. This corresponds to the part of the scene selected as the most salient by humans [21].

$$s(\mathbf{x}) = \sum_{\sigma \in \Sigma} \sum_{\mathbf{x}' \in N(\mathbf{x})} \|\mathbf{M}^{\sigma}[\mathbf{f}^{\sigma}(\mathbf{x}) - \mathbf{f}^{\sigma}(\mathbf{x}')]\|, \tag{8}$$

where $\mathbf{f}^{\sigma}$ denotes the Gaussian smoothed color image at scale σ, and $\Sigma = \{1,2,4,6,8,10,12,14\}$. $N(\mathbf{x})$ is a 9×9 neighborhood window. $\mathbf{M}^{\sigma}$ is the transformation matrix computed from Gaussian derivatives of scale σ and can be any of the three methods mentioned before: $\mathbf{M}_l^c$, $\mathbf{M}_g^c$, or $\mathbf{M}_o^c$. Note that leaving out $\mathbf{M}$ from Eq. (8) results in the multiscale contrast approach proposed by Liu *et al.* [21]. An example of a multiscale color saliency map is given in Fig. 2. The edges of the salient pastry are considered more salient by the multiscale color saliency map.

3. PSYCHOPHYSICAL EVALUATION OF BOTTOM-UP SALIENCY

In this section, a psychophysical experiment is presented that allows us to quantify the accuracy with which information theory (represented by the computational saliency measures) predicts bottom-up saliency in humans. Here we regard saliency as the degree with which an item stands out from its surroundings.

Different psychophysical methods for measuring bottom-up saliency exist, like eye tracking and fixation analysis (e.g., [22]), reaction time analysis (e.g., [2]), and target location (e.g., [23]). We decided to measure the relative saliency of simple center-surround test patterns that we synthesized with specified distributions of chromatic transitions (Fig. 3). These patterns have been designed with the aim that they do not show any familiar object or shape and thus avoid possible effects of cognitive top-down mechanisms. Two center-surround patterns, differing only in chromatic distribution of the centers, are shown side by side to the observers. They decide which of the two centers stands out most from the common surround, i.e., has a higher relative saliency [24].

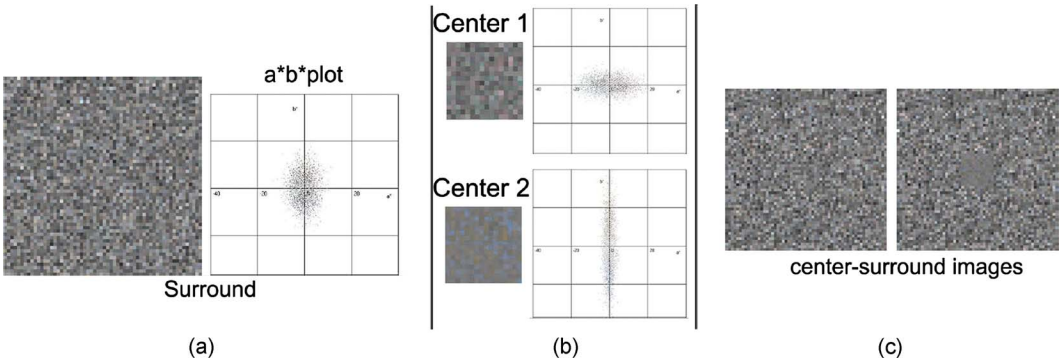

Fig. 3. (Color online) (a) Example of a synthetic image with specified distribution of color transitions in CIELab color space, which forms the surround. (b) Two different transformations of the color distribution shown in (a) form two different centers. (c) Layout of the psychophysical experiment, showing two center-surround color patterns side by side. The surrounds are identical, the centers are different. Subjects indicate which of the two centers stands out most from the surround.

The color patterns are defined in the CIELAB color space [25] with the goal to specify colors in terms of a perceptual space and enable comparison of the results with other studies [26]. Given a certain distribution of L*a*b* values in the surround, the saliency of the center will depend on the difference between the L*a*b* distribution of the center and that of the surround. The more the two distributions differ, the higher the chromatic contrast, which usually results in higher experimental measurements of saliency. The aim is to determine how strong this saliency depends on the underlying L*a*b* distributions (and associated edge transition distributions). We therefore transform these distributions in a systematic manner. For example, under one condition we let the distribution of the surround correspond to the distribution of edge transitions as determined from the COREL image data set.

The data from our observers is compared with predictions on the basis of our models of computational saliency. The latter are applied to the same center-surround images as presented to our observers. A prediction of the center-surround pattern with the higher relative saliency is simply obtained by comparison of the model outputs for the two center-surround images.

A. Method

1. Subjects

Five men and three women (ages ranging from 22 to 29) participated in our experiment. They had normal or corrected-to-normal acuity and normal color vision as confirmed by testing on the HRR pseudoisochromatic plates (4th edition). Subjects were unaware of the purpose of the experiment.

2. Apparatus

Stimuli were presented on a calibrated LCD monitor (Eizo, ColorEdge CG211) operating at 1600×1200 pixels (0.27 mm dot pitch) and 24-bit color resolution. Using a spectrophotometer (GretagMacbeth, Eye-one) the monitor was calibrated to a D65 white point of 80 cd/m^2, with gamma 2.2 for each of the three color primaries. CIE 1931 x,y chromaticity coordinates of the primaries were $(x,y)=(0.638,0.322)$ for red, (0.299,0.611) for green, and (0.145,0.058) for blue, respectively, closely approximating the sRGB standard monitor profile [27]. The spatial uniformity of the display, measured relative to the center of the monitor, was $\Delta E^*_{ab}<1.5$ according to the manufacturer's calibration certificates. This type of display was shown to provide color reproduction errors of the order of 1 just noticeable difference [26], accurate enough for the type of experiment described in this paper.

3. Stimuli and Design

Figure 3(c) shows the layout of the experiment. Two center-surround patterns, differing only in the center, are shown side by side. We used four surrounds as listed in Table 1. The statistics (color edge distribution) of the first surround, labeled S_L, correspond with the statistics of natural images contained in the COREL image data set. From Fig. 1 it is observed that for the COREL data set we have five times more transitions (edges) in intensity than transitions in RG and BY. In L*a*b* space this corresponds to $\sigma_{L_{Corel}}=54$, $\sigma_{b_{Corel}}=27$, and $\sigma_{a_{Corel}}=16$.

The second and third surround, S_a and S_b, were obtained by switching the distributions along the L^* and a^* axes, and along the L^* and b^* axes, respectively. The last surround, S_{eq}, has equal distributions (i.e., amounts of information) in all three directions.

To create the centers in the center-surround patterns, we transform the distribution of the surround by multiplying σ for each axis by a certain value. Each surround listed in Table 1 was combined with 13 different center distributions. These center distributions were obtained by applying the transformation

$$\begin{pmatrix}\sigma'_L\\ \sigma'_a\\ \sigma'_b\end{pmatrix}=\begin{pmatrix}\alpha & 0 & 0\\ 0 & \beta & 0\\ 0 & 0 & \gamma\end{pmatrix}\begin{pmatrix}\sigma_L\\ \sigma_a\\ \sigma_b\end{pmatrix}. \quad (9)$$

The first transformation, labeled C_1, had values $\alpha_0=(\sigma_{a_{Corel}})^{-1}$, $\beta_0=(\sigma_{b_{Corel}})^{-1}$, and $\gamma_0=(\sigma_{L_{Corel}})^{-1}$. This transformation is predicted by our computational saliency measure $\mathbf{M}_l^c$ (computational local transformation, which is here fixed to L*a*b* space) as the most salient between all possible transformations. Five more center patches (C_2–C_6) are generated with α_0, β_0, and γ_0 interchanged. Centers C_L, C_a, and C_b were created by maximizing the energy of the axis indicated by the subscript, while the energy in the remaining two axes are equal. Centers C_{La}, C_{Lb}, and C_{ab} were created by maximizing the energy of two axes (indicated by the subscripts). Finally, center C_{eq}

Table 1. Surrounds with Systematic Changes in the Standard Deviations (σ) Along the L*, a*, and b* Axes of Perceptual Color Space[a]

Surround	σ_L	σ_a	σ_b
S_L	$\sigma_{L_{Corel}}$	$\sigma_{a_{Corel}}$	$\sigma_{b_{Corel}}$
S_a	$\sigma_{a_{Corel}}$	$\sigma_{L_{Corel}}$	$\sigma_{b_{Corel}}$
S_b	$\sigma_{b_{Corel}}$	$\sigma_{a_{Corel}}$	$\sigma_{L_{Corel}}$
S_{eq}	$\sigma_{L_{eq}}$	$\sigma_{a_{eq}}$	$\sigma_{b_{eq}}$

[a]The statistics of the first surround (S_L) comply with the energy distributions of natural images contained in the COREL image data set. The last surround (S_{eq}) has equal amounts of energy in the three directions ($\sigma_{L_{eq}}=\sigma_{a_{eq}}=\sigma_{b_{eq}}$).

was obtained by having the same amounts of energy in all three axes.

Summarizing, for each of the four surrounds (backgrounds) we generate 13 different centers (foregrounds) by 13 transformations of the surround statistics. One of these centers is predicted from the computational saliency measure as the most salient. The question is whether human observers also find this center to be the most salient. If so, this means that information theory is a valid underlying mechanism for bottom-up saliency.

4. Procedure

After passing the color vision test, the subjects were seated at a 50 cm viewing distance from the LCD monitor. In each trial, they indicated (by pressing keys on the keyboard) which of the two centers (left or right) stood out most from the surround, i.e., had the highest salience. They were encouraged to make a decision although they could also indicate that the two centers were equally salient.

B. Experimental Results

In each trial, a subject indicated which of the two centers was most salient. Each center was in competition with the 12 others just once. From these trials we compute the relative saliency of each center from the number of times the center was selected as the most salient. In Fig. 4, the relative saliency is shown obtained for all surrounds, in descending order of saliency. Error bars indicate the standard error of the mean, obtained by averaging over the eight observers. The data did not indicate one or more of the observers to be an outlier.

Regarding surround S_L, Fig. 4 shows that center C_{ab} has the highest relative saliency. This is the expected result because S_L has the largest variance in the L* dimension and C_{ab} has a color edge distribution amplified along both the a* and b* dimensions, at the cost of reducing energy in the intensity edge (L*) distribution. Thus, center C_{ab} looks more strongly colored but with less luminance contrast, which is highly salient in the S_L surround. In contrast, the least salient center (C_L) has increased the energy in the intensity edges, at the cost of reducing energy in the color edges. However, since the surround S_L already has a distribution that dominates in the intensity edges, the extra amplification in intensity edges does not result in visual saliency, as predicted for our saliency measure.

Surround S_a was created by rotating the axes of edge distributions such that the largest variance coincided with the a* axis of CIELAB space. This results in an increased edge distribution along the red–green axis of color space, i.e., the colors along the red–green axis become more saturated, at the cost of a decreased edge intensity. Figure 4 shows that for this background the most salient center is C_b and the least salient center is C_6. Note that there is no significant difference between the saliency of C_4 and C_6. Center C_b is most salient because it is amplified along the b* axis (the blue–yellow axis in color space), which is orthogonal to the amplified a* axis of the surround, at the cost of reduced energy in the b* and L* axes. Blue–yellow edges with decreased intensity edges are salient in a dominant red–green edge distribution. Center C_6 and C_4 are least salient in surround S_a because their

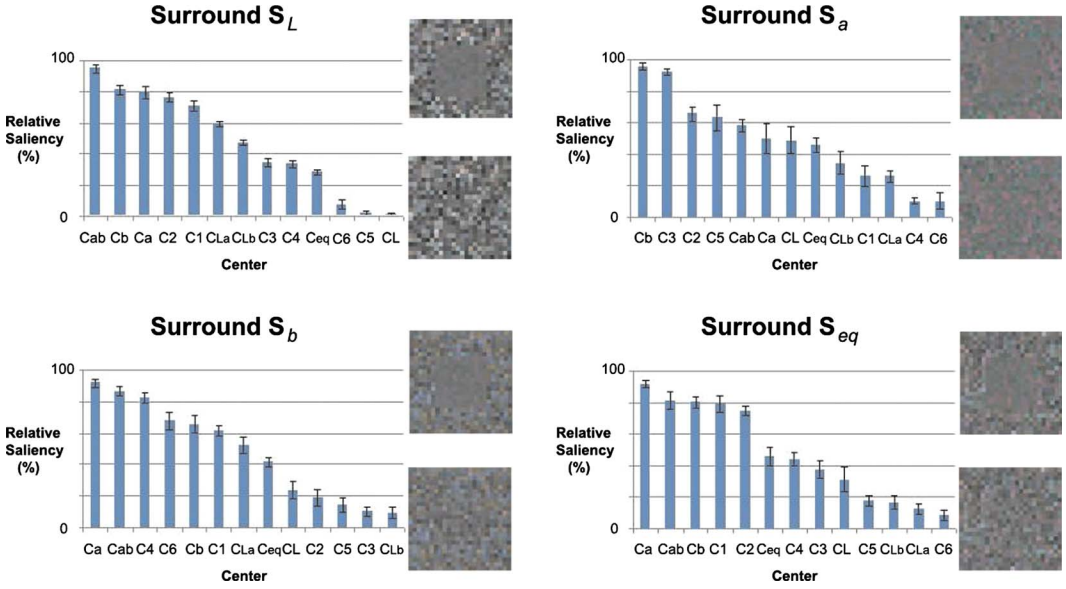

Fig. 4. (Color online) Relative saliency (in descending order) of the 13 centers for the surrounds S_L, S_a, S_b, and S_{eq} averaged over observers. Error bars represent the standard error of the mean. The images on the right-hand side show the most salient (top) and least salient (bottom) centers, in a small portion of the surround.

γ-coefficient in the saliency transformation equals α_0, which is the largest ($\alpha_0 > \beta_0 > \gamma_0$). Thus, intensity edges are amplified most but do not show up as salient in the dominating red–green surround.

The results for surround S_b are described in a similar way as for S_a, but with the role of the red–green and yellow–blue axes interchanged. Thus, in short, C_a is most salient because it has amplified red–green edges (at the cost of blue–yellow and intensity), which stands out from the dominating blue–yellow surround.

The surround S_{eq} is characterized by equal edge distributions along the L*, a*, and b* axes of CIELAB color space. Center C_a apparently is most salient, followed by C_{ab} and C_b, which are all chromatic transformations. The least salient centers are all intensity amplifications. This is an important result: when the edge distributions in the three axes of color space are equal, the most salient change to that distribution is a chromatic one, i.e., an increase of edges along the a* or b* axis, or both, at the cost of a decrease of energy in intensity edges.

With respect to the natural surround S_L (the surround having the statistics of the COREL data set) there remains one important question. Why was center C_1 not the most salient one? We recall that C_1 was expected to be most salient from a computational point of view. Figure 4 shows that C_1 and C_2 are not significantly different and have a higher relative saliency than C_3 and C_4 and C_5 and C_6. Thus, C_1 has indeed the highest saliency with respect to the group of centers C_1 to C_6, as predicted by information theory. Nevertheless, C_1 is still outperformed by the chromatic transformations C_{ab}, C_a, and C_b. The reason for this is that the latter transformations have maximized energy in one or two axes with a total amount of energy that exceeded the total energy of the transformation of C_1. Therefore, the chromatic transformations C_{ab}, C_a, and C_b are more salient than C_1. In conclusion, for the center distributions C_1 to C_6 having the same total energy (information content) in the edge distributions, information theory correctly predicts C_1 to have a high relative saliency in the natural surround S_L.

C. Comparison Between Computational Saliency Models and Psychophysics

Here we compare the performance of the different computational saliency models detailed in Section 2, namely, global opponent color-space saliency (M_o^c) and local computational (M_o^l). Additionally we compare these methods against the Itti model [16] and RGB edges on predicting the human response (the selection of the most salient center) in our psychophysical experiment. We apply a transformation to the matrices obtained in Section 2 to convert them to L*a*b* space. For each subject ($s=1\ldots8$) and each computational model ($m=1\ldots5$), we computed the overall correspondence between the subject's selection and the model's selection of the most salient center. This correspondence $\mathrm{Cor}(s,m)$ is a value between 0 and 100 and is computed as follows:

$$\mathrm{Cor}(s,m) = 100\frac{\sum_{i=1}^{468} a_i}{468}, \tag{10}$$

where a_i denotes—per trial i—the agreement (either 0 or 1) between model and subject. Figure 5 shows the correspondence for the four computational models. Trials in which subjects could not decide on the most salient center are left out of the computation.

It is clear from Fig. 5 that global opponent-space saliency ($\mathbf{M}_o^c$ with 74.75% correspondence) outperforms the other models ($\mathbf{M}_l^c$ and Itti having 62.56% and 57.70% correspondence, respectively). At the 95% confidence level significant differences exist between global ($\mathbf{M}_o^c$) and local computational saliency ($\mathbf{M}_l^c$) ($p=1.1\mathrm{E}{-4}$), between ($\mathbf{M}_l^c$) and Itti ($p=1.2\mathrm{E}{-3}$) and Itti and RGB ($p=1.8\mathrm{E}{-15}$). We also computed the interobserver agreement using Eq. (10) but with a_i replaced by w_i, where w_i represents the fraction (between 0 and 1) of subjects that gave the same response in each trial i. Thus, if 6 of the 8 subjects selected the same center, $w_i=6/8$. This resulted in an observer agreement of 86.8%. In conclusion, our computational saliency methods (both local and global) are significantly better at predicting human saliency than the Itti and Koch model, as shown in Fig. 5.

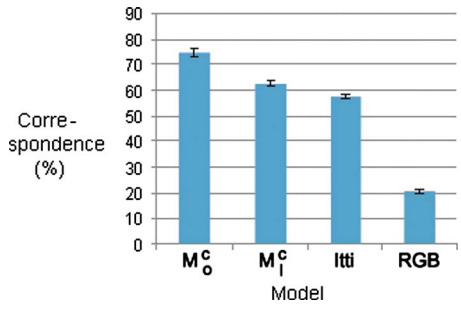

Fig. 5. (Color online) Correspondence as computed with Eq. (10) between computational saliency models. The different computational models are sorted on descending correspondence. Error bars indicate standard error of the mean (eight subjects).

4. EVALUATION ON HIGH-LEVEL SALIENCY

In this section, we validate computational saliency for the task of salient object detection in real-world images and compare it with existing methods. We call this task *high-level saliency* to distinguish it from saliency detection that is purely bottom-up. Bottom-up saliency was investigated in the previous section, where patterns were used that do not show any familiar object or shape avoiding possible effects of cognitive top-down mechanisms. In high-level saliency tasks, top-down information such as image semantics, is also considered.

A. High-Level Saliency Data Set

To compare our computational saliency methods with high-level saliency, we use a large-scale image data set of human labeled salient objects [21]. Example images of this data set are shown in Fig. 6.

The data set contains a large number of high-quality images obtained from different sources such as image forums and image search engines. Images all contain a single salient object or a distinctive foreground object. For each image, users drew a rectangle enclosing the most salient object in the image. We use the set B consisting of

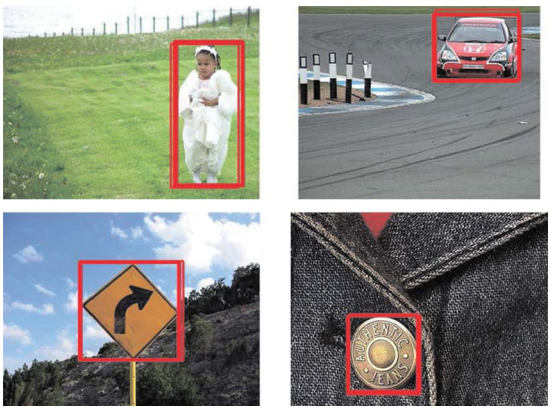

Fig. 6. (Color online) Labeled images from image set B consisting of 5000 images that were labeled by 9 users obtained from [21].

5000 images that were labeled by 9 users[21]. Foreground pixels are those pixels that are considered to be foreground by a majority of the users. Then, this set is divided into 10 subsets of 500 images each $(B_1,\ldots,B_{10})$. We use the 500 images in B_1 for training and the rest of the 4500 images for testing.

B. Assessing Computational Saliency Transformations

In Subsection 2.A, we proposed different computational saliency measures based on information theory. Here, we want to verify to what extent the transformation given by information theory corresponds to the optimal linear transformation possible on a given data set. To this end, we compute the optimal transformation, called $\mathbf{M}_o^h$, on the labeled data set.

To determine the success of each transformation, the precision index [21] is used, which is computed as follows. An image is divided into a foreground region f^i and a background b^i, where i is the image index. Let $f_{\mathbf{M}}^i$ be the summed saliency of the foreground for a certain saliency transformation $\mathbf{M}$. Let $b_{\mathbf{M}}^i$ denote the same for the background. Further, let $A(f^i)$ and $A(b^i)$ denote the area of the foreground and background, respectively. The confidence measure used is the precision index P_Λ^i:

$$P_{\mathbf{M}}^i = \frac{A(b^i) f_{\mathbf{M}}^i}{A(f^i) b_{\mathbf{M}}^i}. \tag{11}$$

In other words, $P_{\mathbf{M}}^i$ provides the likelihood to select from the image a location that is within the salient bounding box [21]. A higher precision index therefore corresponds to a better saliency measure.

To reduce the set of possible transformations, we restrict the transformation to the opponent color space transformation. We define $\mathbf{M}_o^h$, called *learned saliency*, as the transformation that maximizes $P_{\mathbf{M}}^i$ by varying the parameters $\Lambda=\mathrm{diag}(\alpha,\beta,\gamma)$. An exhaustive search based on all training set images is performed. Hence, $\mathbf{M}_o^h$ is the transformation that obtains the maximum correspondence (given the opponent transformation) to the human assessments of object saliency.

A high correlation is expected between $\mathbf{M}_o^h$ and the global opponent color-space saliency transformation $\mathbf{M}_o^c$. Table 2 summarizes the results of the computational saliency and the learned saliency measures in terms of the precision index.

When comparing the learned saliency measure with the computational saliency measure, it can be inferred that the results obtained by the computational approach are indeed in agreement with the best possible transformation, that is, the learned saliency measure. In both cases, γ is a fairly small value. This is because there is a high amount of achromatic transitions in the images as opposed to chromatic ones. Hence, these transitions are less informative, as predicted by the computational saliency measure. To obtain a proper saliency map, the weights of these transitions should be decreased. Furthermore, α and β values are larger and close to each other in both cases. It is also interesting to note that local saliency outperforms global saliency. This is because local saliency can adapt its transformation for each individual image.

To quantitatively show the resemblance of the saliency maps computed by the computational and learned saliency, we have calculated the intersection of the normalized saliency maps. The intersection is computed by taking the inner product between two saliency maps (the saliency maps are first transformed to vectors). The averaged score over all images reaches 97.43%. For comparison, the overlapping between learned saliency and saliency based on RGB edges (without additional transformation) is only 83.11%. A qualitative comparison between computational and learned saliency is depicted in Fig. 7.

In conclusion, the relevance of using information theory as a saliency processing model from a computational point of view has been demonstrated.

C. Evaluation of Computational Saliency Methods on High-Level Saliency

Here, we evaluate the different computational saliency models proposed in this paper and compare them with the Itti and Koch saliency method [16]. The evaluation is performed on the real-world image data set [21]. To evaluate saliency methods, we use the hit and miss index (a comparison measure commonly used in the literature): if the maximum of the saliency map falls inside the original rectangle, we have a hit, otherwise, a miss is registered. Note that for each image the size of the foreground (rectangle) is given.

We evaluate the learned saliency ($\mathbf{M}_o^h$), computational global saliency ($\mathbf{M}_o^c$), and computational local saliency ($\mathbf{M}_l^c$). In addition to these transformations, we also show results obtained with the multiscale computational local transformation ($\mathbf{M}_l^{\sigma c}$), the RGB edges without any trans-

Table 2. Results Obtained for Learned Global Saliency Measure $\mathbf{M}_o^h$, Computational Global Saliency $\mathbf{M}_o^c$, and Computational Local Saliency $\mathbf{M}_l^{ca}$

Measure	α	β	γ	$P_{\mathbf{M}}^i$
$\mathbf{M}_o^h$	0.65	0.34	0.01	0.49
$\mathbf{M}_o^c$	0.53	0.43	0.04	0.45
$\mathbf{M}_l^c$	Image dep.	Image dep.	Image dep.	0.51

[a]The fourth column shows the average precision score.

Fig. 7. (Color online) Color saliency example. First row: original image. Second row: RGB edges. Third row: computational global saliency $\mathbf{M}_o^c$ (see Table 2). Fourth row: $\mathbf{M}_o^h$ (see Table 2). The overlap between learned and computational maps over all images reaches 97.43%, whereas the overlapping with the RGB edges is 83.11%.

formation (RGBe), the Itti saliency method (Itti), and a random selection of the most salient location (Random). Table 3 summarizes the results obtained.

From this table, it can be concluded that the results obtained with multiscale contrast are better than others. Compared with the local saliency model, the hit index increases from 89.6 to 95.2. Furthermore, using locally induced saliency provides better performance than computing the color transformation based on color edges extracted from the whole image data set. As expected, locally computing the transformation adapts better to the edge distribution for each image. Furthermore, the results show that locally induced saliency with multiscale contrast provides the best performance.

Table 3. Hit and Miss Values Obtained in the Test Set for All Proposed Saliency Transformations as Well as for RGB Edges, Itti Saliency Measure [16], and a Random Selection of the Most Salient Location

Transformation	Hit	Miss
Global learned ($\mathbf{M}_o^h$)	87.1	12.9
Global computational ($\mathbf{M}_o^c$)	87.9	12.1
Local computational ($\mathbf{M}_l^c$)	89.6	10.4
Local multiscale computational ($\mathbf{M}_l^{\sigma c}$)	95.2	4.8
RGB edges	81.4	18.6
Itti	88.2	11.8
Random	72.8	27.2

5. CONCLUSIONS

In this paper, different computational methods are proposed to compute color edge saliency based on the information content of color edges. A comparison has been done between these computational models and human perception. First, a psychophysical experiment has been conducted using patterns without semantic image cues (bottom-up saliency task). Second, a computational experiment has been done focusing on the task of salient object detection including object and scene semantics (high-level saliency).

From the psychophysical experiment, the relevance of using information theory as a saliency processing model from a human-perception point of view has been demonstrated. It can be derived that for a uniformly distributed background, humans are more sensitive to chromatic

changes than luminance variations. Furthermore, it is shown that the proposed method performs significantly better at predicting saliency (with a human-method correspondence of 74.75% and an observer agreement of 86.8%) than state-of-the-art models.

From the computational experiment, it has been shown that the use of information theory as a saliency processing model is also valid from a computational point of view. The results obtained from a large-scale data set confirm that an early fusion of these features yields an improvement on the prediction of saliency. Furthermore, it is shown that the proposed computational methods provide accurate performance to compute visual saliency with a hit rate up to 95.2%.

ACKNOWLEDGMENTS

This work has been partially supported by projects TIN2004-02970, TIN2007-64577, TIN2009-14173, and Consolider-Ingenio 2010 CSD2007-00018 of the Spanish MEC (Ministery of Science), ERG-TS-VICI-224737 (European Re-integration Grants), the Ramon y Cajal Program, and the EU-NEST PERCEPT.

REFERENCES

1. E. Titchener, *Lectures on the Elementary Psychology of Feeling and Attention* (Adamant Media Corporation, 2005).
2. A. Koene and L. Zhaoping, "Feature-specific interactions in salience from combined feature contrasts: evidence for a bottom-up saliency map in V1," J. Vision **7**(7), 6 (2007).
3. A. Torralba, A. Oliva, M. Castelhano, and J. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search," Psychol. Rev. **113**, 766–786 (2006).
4. T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *European Conference on Computer Vision*, 2004, pp. 228–241.
5. G. Fritz, C. Seifert, L. Paletta, and H. Bischof, "Attentive object detection using an information theoretic saliency measure," in *Attention and Performance in Computational Vision: Second International Workshop, WAPCV 2004*, Revised Selected Papers, 2005, pp. 29–41.
6. M. Mancas, B. Unay, B. Gosselin, and D. Macq, "Computational attention for defect localisation," in *Proceedings of ICVS Workshop on Computational Attention & Applications*, 2007.
7. N. Bruce and J. Tsotsos, "Saliency based on information maximization," Adv. Neural Inf. Process. Syst. **18**, 155–162 (2006).
8. D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," Adv. Neural Inf. Process. Syst. **17**, 481–488 (2005).
9. D. Gao, V. Mahadevan, and N. Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency," J. Vision **8**(7), 13 (2008).
10. L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell, "SUN: a Bayesian framework for saliency using natural statistics," J. Vision **8**(7), 32 (2008).
11. D. Gao and J. Zhou, "Adaptive background estimation for real-time traffic monitoring," in *2001 IEEE Intelligent Transportation Systems*, 2001, pp. 330–333.
12. T. Jost, N. Ouerhani, R. Wartburg, R. Muri, and H. Hugli, "Assessing the contribution of color in visual attention," Comput. Vis. Image Underst. **100**, 107–123 (2005).
13. F. Wichmann, L. Sharpe, and K. Gegenfurtner, "The contributions of color to recognition memory for natural scenes," Learn. Memory **28**, 509–520 (2002).
14. J. Wolfe and T. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" Nat. Rev. Neurosci. **5**, 495–501 (2004).
15. H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 222–228.
16. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998).
17. J. Wolfe, "Guided Search 4.0: current progress with a model of visual search," in *Integrated Models of Cognitive Systems*, 2007, pp. 99–119.
18. Z. Li, "A saliency map in primary visual cortex," Trends Cogn. Sci. **6**, 9–16 (2002).
19. J. Krummenacher, H. J. Muller, and D. Heller, "Visual search for dimensionally redundant pop-out targets: evidence for parallel-coactive processing of dimensions," Percept. Psychophys. **63**, 901–917 (2001).
20. J. van de Weijer, T. Gevers, and A. Bagdanov, "Boosting color saliency in image feature detection," IEEE Trans. Pattern Anal. Mach. Intell. **28**, 150–156 (2006).
21. T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," in *Proceedings of IEEE Computer Society Conference on Computer and Vision Pattern Recognition*, 2007, pp. 1–8.
22. D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," Vision Res. **42**, 107–123 (2002).
23. M. J. Wright, "Saliency predicts change detection in pictures of natural scenes," Spatial Vis. **18**, 413–430 (2005).
24. H. C. Nothdurft, "Salience from feature contrast: additivity across dimensions," Vision Res. **40**, 1183–1201 (2000).
25. G. Wyszecki and W. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., 2000, p. 968.
26. M. Lucassen, P. Bijl, and J. Roelofsen, "The perception of static colored noise: detection and masking described by CIE94," Color Res. Appl. **33**, 178–191 (2008).
27. M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "A standard default color space for the Internet sRGB," *Microsoft and Hewlett-Packard Joint Report*, Version 1, 1996.