Received: 19 December 2017 | Revised: 22 March 2018 | Accepted: 23 March 2018

DOI: 10.1002/col.22235

RESEARCH ARTICLE

WILEY

Heuristic analysis influence of saliency in the color diversity of natural images

Juan Luis Nieves | **Javier Romero**

Department of Optics, Faculty of Sciences, University of Granada, Granada 18071, Spain

Correspondence
Juan Luis Nieves, Department of Optics, Faculty of Sciences, University of Granada, Granada 18071, Spain.
Email: jnieves@ugr.es

Funding information
Ministry of Economy and Competitiveness of Spain, Grant/Award Number: DPI2015-64571-R

Abstract

The estimation of chromatic diversity of natural images is commonly quantified through the computation of the number of discernible colors and has received much attention because of the different implications it has. However, the relationship between that number and the number of colors that really attracts the attention from an observer is still not clear and has been given little attention. New concepts about *salient discernible colors*-the salient chromatic diversity of images- and *remarkable salient colors*-connected colors in the same salient image area-are introduced as opposed to the classical number of discernible object colors, which is usually evaluated for the global image without differentiating between probable attended and non-attended image regions. We have used different well-known saliency models to locate the salient regions in the scenes and have heuristically studied the extent to which those models preserve the chromatic diversity of natural images. Based on a bottom-up approach, a reduction of around 40%-55% in the number of discernible colors were obtained, and not all saliency algorithms preserved a uniform sampling of the original color gamut. Thus, our results suggests that particularly the graph-based visual saliency model got good low dissimilarity values in comparison with other approaches that put emphasis solely on color as the main low-level feature. Furthermore, we have introduced a quantification scheme of the average number of remarkable salient colors appearing in the images, and have proved how the heuristic-based analysis of salient image areas can be used to create segmented images automatically according to their salient chromatic diversity.

KEYWORDS

color, color identification, color imaging, visual saliency

1 | INTRODUCTION

Color in nature and natural image statistics have garnered lot of attention during the last years supporting the idea that human visual system evolved to be adapted to the natural environment.[1,2] That efficient coding hypothesis is one of the links between the statistical and the visual properties of the visible surroundings. This means that human visual system is able to analyze in an optimal way most of the visual features in the visual field such as the detection of edges in images, the processing of colors and contrasts or the visual motion estimation.

So far, different studies have characterized the properties of the natural images with the aim to understand our visual coding.[3] Among those properties the chromatic diversity has been recently introduced to analyze the color gamut and color volume expanded by natural and artificial images.[4] The estimation of chromatic diversity is quantified through the computation of the number of discernible colors in images and has been used in different studies and applications (e.g., gamut computation in displays, color reproduction and rendering in museums, etc.). The general principle to estimate the number of discernible colors is to segment the color space in just noticeable sub-volumes and to count the

number of these volumes containing the color representation of 1 pixel minimum. Being a classical problem with a long history, it can be stated that there are around 2 million distinguishable colors under illuminant D65 (see Masaoka and Berns[5] for a further review). Nevertheless the relationship between the number of discernible colors and the number of colors that really attracts the attention from an observer is still not clear and demands further analysis.

Therefore it is of great interest, from both theoretically and practical point of views, to look for appropriate answers about what determines in the end where and why an observer points his/her gaze to particular locations in a scene. What is clear is that some areas in an image can attract the visual attention and the point of gaze of an observer scanning the visual field. In doing so, it is said that those regions show good saliency, that is, specific low-level visual features are attracting the observers' interest,[6] and thus the saliency map is a biologically plausible model for bottom-up overt attention proposed by Koch and Ullman.[7] Their definition of saliency relied on center-surround principles considering that points in the visual scene are salient if they differ from their neighbors. There are many features characterizing a visual scene, among which we could cite edges, contrast, luminance and color as main visual features defined at different scales. Classical stimulus-driven and bottom-up visual models get relatively good results when they used those features to localize the highly salient features in a scene, both for natural and artificial images. Latterly, including task-dependent constraints within the saliency algorithms, operating at higher levels, improves quantitatively the derived salient maps.[8] Those kinds of models use a prior knowledge to get the visual attention, being a top-down model. The main drawback is that a testing step is necessary before getting and predicting the final salient object map, which is not always possible. From the computational point of view, we should add the hybrid models that combine the advantages and simplicity of the bottom-up models with the goal-task efficient constraint imposed by the top-down models.[9]

Many projects have been developed in recent years trying to find algorithms and models to characterize where humans look, that is, the saliency map, in a scene base on a reduced number of visual features. Eye tracking systems are usually employed to record observers' gaze paths as they view a collection of images. After discarding saccades fixation locations and the corresponding fixation map can be obtained.[10] In the review of the model by Itti et al.,[11] the visual saliency maps are topographical codifications of fixation position in visual search over the entire scene based on different image features such as luminance, orientation or color. The algorithm is based on bottom-up scene-based properties by selecting pre-attentively computed simple features (color, intensity, and orientation) and combines all of them into a conspicuity map for each channel. The combination of the conspicuity maps into a saliency map produces the serial selection of salient locations in the image.

In addition, as established by Tatler et al.[12] there are differences between visual features in attended and non-attended spatial locations in an image. To be specific, those differences are determined by various contrasts, luminances, chromaticity, energy, and orientation. Nevertheless those findings have been cast doubt on by Braddely and Tatler[13] who found that fixation map is dominated by the high-frequency edges; by so doing the authors argue that contrast does not contribute to saliency and that the other features are "behaviorally irrelevant." That being said however, other authors based salient detection on chromatic content of natural scenes. In the Tian et al.[14] a color saliency model is introduced based on a hue–saturation–intensity representation of images, which is correlated to 5 features (contrast of hue, contrast of saturation, contrast of intensity, dominance of warm colors, and dominance of brightness and saturation) before getting the final saliency map as a weighted linear combination of those feature maps. The model performs well in detecting color regions in natural scenes and their results are very promising for color scene understanding.

If the reader is interested in full details about the performance of the above cited saliency models and other computational approaches, Bylinskiia et al.,[15] Borji and Itti,[16] and Judd at al.,[17] have created excellent taxonomies of visual attention literature and benchmark data sets comparing performances to baseline models under different metrics.

2 | AIM OF THE WORK

There are several studies analyzing how color influences visual attention and the fixations locations in natural scenes.[12,18,19] In this study, we heuristically examine the effects of visual saliency in the estimation of the number of discernible colors in natural scenes and explore the relationship and consistency between that number and the salient colors. Different well-known bottom-up saliency models created from a wide variety of different approaches have been used to locate salient regions in the scenes where the number of discernible colors will be computed.

A new concept about "salient discernible colors," which are defined as the discernible colors that are salient—the salient chromatic diversity—is introduced as opposed to the classical number of discernible object colors, which is evaluated for the global image without differentiating between probable attended and non-attended image regions. It should be made clear that this article does not deal with the issue about the computation of the maximum number of mutually distinguishable colors. The initial assumption is made that salient regions will attract observer attention and could affect the estimation of the total number of distinguishable colors perceived in an image. A final segmentation and labeling

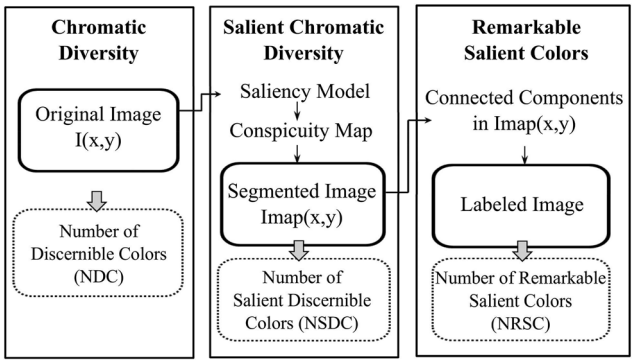

FIGURE 1 Steps for computing and estimating the Number of Discernible Colors (NDC), the number of Salient Discernible Colors (NSDC), and the Number of Remarkable Salient Colors (NRSC) in images

step is also proposed to analyze which colors will be naturally attracting observers' attention in natural images, and the term remarkable salient colors is used to label and compute those colors in each image.

3 | METHODS

The flowchart in Figure 1 summarizes all steps for computing and estimating the Number of Discernible Colors (NDC), the number of Salient Discernible Colors (NSDC), and the Number of Remarkable Salient Colors (NRSC) in images.

3.1 | Images

We have used a set of natural images that were collected from 2 different image data sets. First, we used 24 hyperspectral natural images-denoted as "Nascimento's set" in this article-that were rendered under a daylight illumination of 6489 K correlated color temperature (quite close to the standard 6500 K daylight). Size of images were 500 × 500 pixels, all images at full resolution are available online at http://online.uminho.pt/pessoas/smcn/hsi_2002/hsi_2002.html; additional details about spectral image acquisition, assumptions and approximations to get the final spectral reflectance $r(x,y;\lambda)$ at a pixel (x,y) in the image can be found in Nascimento et al.[20] By selecting this data set you can completely control the illumination under which images are rendered and thus the color of scenes as viewed by a hypothetically observer. The spectral reflectance of each pixel was converted into spectral radiance in the usual way as the product of each spectral reflectance $r(x,y;\lambda)$ and the spectral power distribution of the illuminant $e(\lambda)$ (assuming constant illumination over the scene), and next data were converted to CIE XYZ tristimulus values by using the CIE 1931 color matching functions,[21]

$$I_k(x,y) = \sum_{\lambda=400}^{700} \bar{q}_k(\lambda) r(x,y;\lambda) e(\lambda) \Delta\lambda \qquad (1)$$

where (x,y) denotes the 2D pixel coordinates in the image, λ is the wavelength of the incident light, k stands for the X, Y, and Z tristimulus values and $\bar{q}_k(\lambda)$ denotes the corresponding color matching functions. The kth image components were converted into the CIE $L^*a^*b^*$ color space for further colorimetric analyses.

Next, we used 150 RGB images-denoted as "Torralba's set"-that were collected from http://saliency.mit.edu/datasets.html.17. We selected only the landscapes images with nature content and rejected seashore, artificial and portrait images; size of images were 400 × 400 pixels. As opposed to the other set of images, illuminant conditions cannot be fully controlled now, and thus we assumed that all RGB images in this data set were rendered under a similar daylight illumination.

Two additional non-natural images were used just for introduction an application of the proposed labeling technique and for comparison purposes. One of the artificial images was extracted also from the Nascimento et al.'s set[20] and the other one corresponded to a painting representing nature (http://art-saus.deviantart.com/art/Magic-Autumn-Colors-322305598).

3.2 | Analysis of the chromatic diversity of scenes

The analysis of the chromatic diversity of scenes is directly linked to the estimation of the number of discernible colors in images. Therefore the number of discernible colors was computed by segmenting the CIE $L^*a^*b^*$ representation of each image in unitary cubes and by counting the number of non-blank cubes. As proposed by Linhares et al.[4] it was assumed that all the colors that were located within the same square, that is, sharing a color difference less or equal to 1 CIE $L^*a^*b^*$ unit, will not be discernible.

3.3 | Computing salient regions in the scenes

All classical bottom-up saliency models and algorithms found in the literature[16,22] share common steps, which can be summarized as follows: selection of different pertinent visual features (such as color, edges, contrast, etc.), computation of a saliency map for every feature (by employing filtering processing that emulates human visual system), and a final integration step of all separate saliency maps into a single output which signals those regions that would attracts observer's gaze and fixation.

During the past decade many saliency models have been proposed, not always being inspired in the neural processing of visual information. In this work we have selected the models from Itti et al.,[11] Harel et al.,[23] Tian et al.,[14] and Canosa.[8] Although there are many models available in the related literature here we decided to choose the aforementioned models because all are easily implemented, all include color as one

of the main low-level features and have been tested in a wide variety of approaches; the selection of the Harel et al.'s was driven by Judd at al.[17] findings that proved that this model performed the best among others. In addition, we have also analyzed other saliency-based models that employ specific color visual features to quantize several representative classes that can be used to differentiate regions in the image. The procedures for saliency map generation are described as follows for each model (refers to the Appendix to get more details about the rationale behind each model):

1. Itti et al.'s model: 3 conspicuity maps are obtained and summed into the final input to the saliency map. Starting from RGB color values of the input image, 3 channels are used computing the L^* intensity image, the red–green and the b^* blue–yellow components. Using steerable filters at each image location local orientations are computed and the model compute center-surround differences to determine contrast. After an across-scale combinations and normalization 3 conspicuity maps are summed into a final single salient map signaling potentially meaningful parts of the scene.
2. Harel et al.'s model, or the so called graph-based visual saliency (GBVS) model: the algorithm highlight salient image locations where the image is informative based on human fixations. At a first stage this model computes feature maps using biologically inspired filters, next it computes an activation map based on spatial locations shown to be unusual in its neighborhood and finally it normalizes it in a way which highlights conspicuity. A Matlab implementation of this model can be found at http://www.vision.caltech.edu/~harel/share/gbvs.php.
3. Tian et al.'s model: the model is designed for detecting natural objects by combining different color contrast-based features. By considering a perceptual object spatially distinguished from the surround through its homogenous color and high contrast, the authors propose 5 features to describe the saliency of color. Three of the features were contrast-based features (contrast of hue, contrast of saturation, and contrast of intensity) and the other 2 features we connected to the dominance of warm colors and the dominance of brightness and saturation. The framework of the model does not include either a multi-scale or an orientation filter bank and thus concentrates only on a local color saliency.
4. Canosa's model: the model introduces a biologically plausible model of selective attention that selects regions of high visual salience in an image and modulates that saliency to locate potential objects in the scene; those potential objects are selected according to their usefulness in the context of an explicit or implicit task. The model is based on the approach taken by Itti and Koch,[6] and Parkhurst et al.[24] in that it uses spatial filters at various resolution levels to detect low-level features of potentially high visual salience. The model uses a proto-object locator to simulate figure/ground segmentation of perceptual organization. At a final stage, the algorithm creates the conspicuity map by combining a color map, an intensity map, an orientation map and the object map.

In addition to those saliency models, we have tested 4 alternative ways of getting the final saliency map by using only color features to compute the most attractive regions:

1. Simple hue–saturation–intensity feature-based model that computes the saliency map based on an equal weight of the hue, saturation and intensity components of each image.
2. A simplified Tian's et al. model that computes only the contrast-based features related to the contrast of hue, contrast of saturation and contrast of intensity.
3. Colorfulness feature-based model: colorfulness is defined as the visual sensation according to which the perceived color of an area appears to be more or less chromatic (see Fairchild[25] on page 87). Within this approach the color salient map is obtained using that standard colorfulness descriptor as the main visual feature.
4. Alternative colorfulness feature-based model that uses the formula proposed by Lübbe[26] to compute the colorfulness attribute.

Figure 2 summarizes all steps starting with the original Input image decomposed into the L^*, a^*, and b^* chromatic components. After applying each saliency model, a 75-percentile was used to threshold the output of each saliency map $I_{map}(x,y)$ and the NSDC was counted within that segmented image; a final output, as a labeled image in terms of its salient colors, was produced. Segmentation could be done by selecting variable percentiles as thresholds. If we include large areas around the maximum of the salient regions certain non-salient parts will be also highlighted (the surround and background for instance). On the contrary, restricting too much the threshold could lead to a reduced number of pixels signaling salient regions. The 75-percentile value was a compromise and an intermediate scale that consistently highlight the most relevant salient regions. In doing so, a clear segmentation of meaningful whole objects is obtained without having to use additional steps.

3.4 | Estimating the remarkable salient regions and colors in the images

At the final stage of our analysis (rightmost column in the Figure 1, and Figure 2 on the bottom part of the flowchart) we estimated the remarkable salient colors appearing in the image by first locating the connected components in each

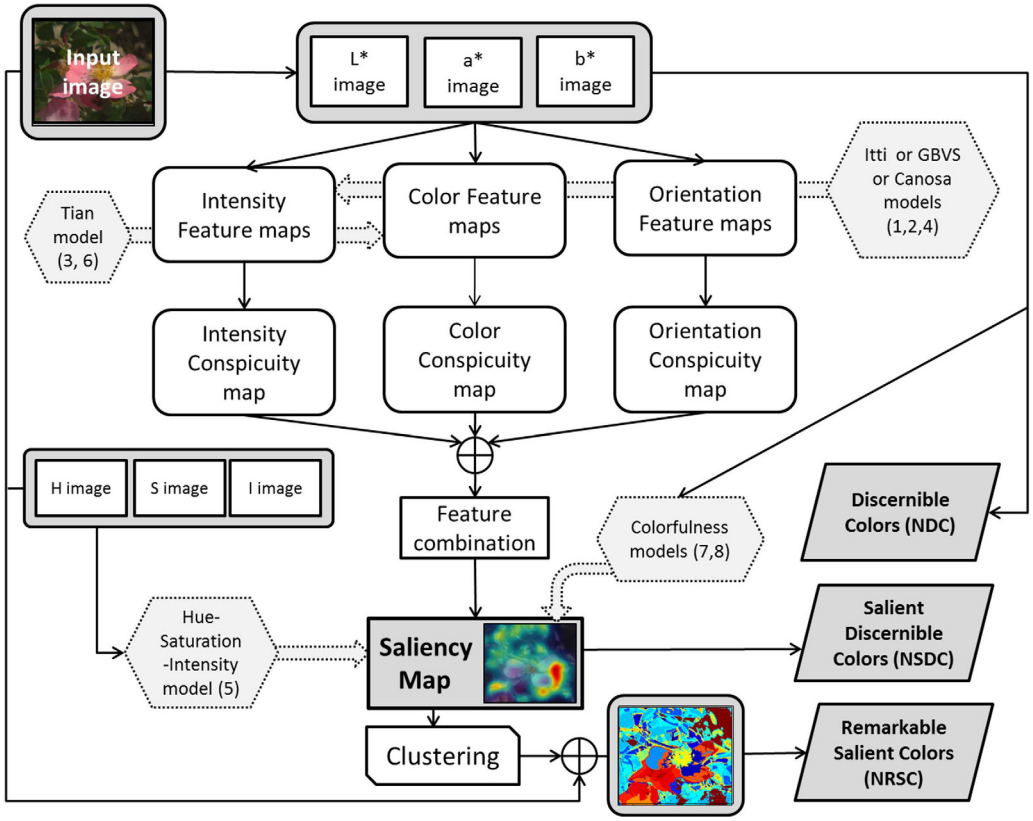

FIGURE 2 Taxonomy and construction of the saliency map and computations of the chromatic diversity of images in terms of the Discernible Colors, Salient Discernible Colors, and Remarkable Salient Colors; every number in rounded brackets is linked to the saliency model explained in Section 3.3

salient map $I_{map}(x,y)$. Connected component labeling is the process of identifying the connected components in an image and assigning each one a unique label. Thus from the saliency map image a binary image $I_{BWmap}(x,y)$ was constructed and the connected components were found in that binary image. A labeled image $I_{labeled}(x,y)$ was then created using a default connectivity of 8 for 2 dimensions, and the remarkable salient colors were assumed to be the center of mass for each labeled region in $I_{labeled}(x,y)$ image.

4 | RESULTS

4.1 | Computing the number of salient discernible colors

We have computed first the number of discernible colors (NDC) following the classical approach, that is, considering the absence of salient features in the scene. This means that the global image is the unique salient source of visual information and thus none saliency model will be introduced in the computation of the NDC. The total NDC obtained for Nascimento's and Torralba's natural images were 20,573 and 22,345, respectively.

In this work we are introducing the concept of "salient discernible colors" to emphasize the influence of color and visual attention clearly in selecting interest areas within an image. Because different visual features can influence and dominate through different image pixels, it is plausible to look for salient colors appearing with different salient weights depending on the scene area scanned by a hypothetical observer. Thus if an observer fixates on specific objects and areas of the image we will assume that only the pixels belonging to those areas will be potential candidates to compute the NDC in the image. By computing the Number of Salient Discernible Colors (NSDC) only discernible colors linked to salient areas of an image will be considered in the estimation of the chromatic diversity. As we will see later on the computation of the salient regions and NSDC will allow also an estimation of the salient colors present in the scene and a segmentation of it according to this criterion.

Figure 3 compares the NDC and the NSDC based on both the different saliency algorithms and the data sets used. Tian et al. model (model 3) and the model using a simple HSI descriptor (model 5) obtained clearly a reduced number of salient discernible colors. The reduction, which is around 82%, is much greater than the obtained with other models, which is around of 40%-55%. In addition, no significant differences were found for the two image data sets with the exception being the NSDC obtained for models 6 and 8.

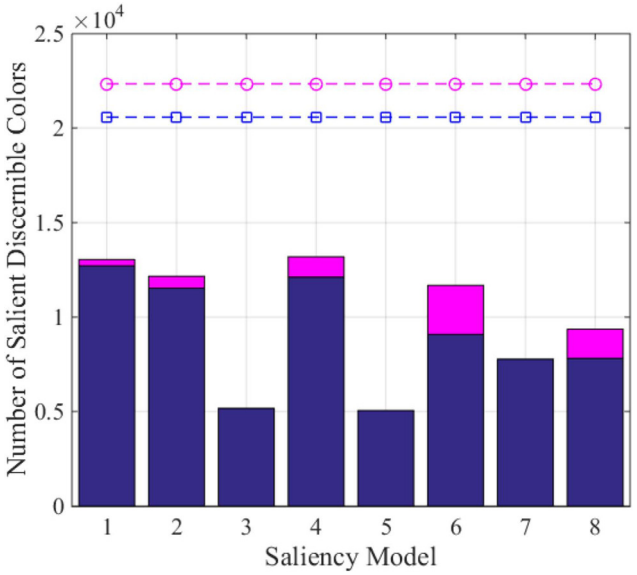

FIGURE 3 Number of Salient Discernible Colors (NSDC) obtained for Nascimento's (blue bars) and Torralba's (purple bars) image data sets and each saliency model; the Number of Discernible Colors (dash lines) are also shown for comparison

Figure 4 shows an example of the obtained results for each model and 1 natural image. As expected, not all saliency models produce the same salient regions. Results suggests that GBVS, the simplified Tian et al., Canosa and the standard colorfulness models locate the flower clearly as one of the salient regions on the image. On the contrary, it is noticeably that the other models select vegetation and the surrounding areas of the flower as the most salient regions, either in the raw map (left column) or in the thresholded saliency map (central column).

Concerning the chromatic distribution of colors (or color gamut), right column in Figure 4 shows the colors in the (a^*, b^*) plane of the whole image (blue markers) and the limited distributions of colors constrained by the salient regions in the image (green markers). In this example, algorithms not capable of selecting the flower as a significant and a salient region show a very restrictive gamut of colors. The distribution of colors derived from Itti et al., GBVS and Canosa models seem to sample uniformly the color gamut from the original image. However Tian et al. and HSI algorithms restricted the chromatic diversity deriving reduced color distributions in comparison with the original image. This tendency was already found in the Figure 3, where those algorithms got a small NSDC suggesting a major contradistinction between the chromatic diversity and the NSDC in natural scenes.

Figure 5 shows the average gamuts obtained for all salient colors predicted for each saliency model in comparison with the original average gamut. It should be first pointed out that chromatic gamut for the Nascimento's set is much more reduced along the negative dimension b^* than the one obtained for the Torralba's set. We have to emphasize that the selected RGB images in the latter case could lead to an unfair comparison between both data sets, with images in the Torralba's set containing more blue sky colors than the Nascimento's set. Results also suggest that Tian et al. (data 3) and HSI models (data 5) removed most of the "reddish" color content (large and positive a^* values) and some extent of the "yellowish" colors (large and positive b^* values), irrespective of the image data sets used. Although Tian et al. model was introduced for characterizing the visual saliency in natural scenes, putting emphasis in the color properties images, is the worst model preserving the chromatic diversity in the images. To quantify that dissimilarity we computed the Hausdorff distance[27] between the original gamut and the corresponding salient gamut. Table 1 summarizes the Hausdorff distances for each of the saliency models used and suggests that the salient gamut obtained from the Canosa algorithm is quite close to the original image gamut for the Nascimento's set (with a Haussdorff distance of

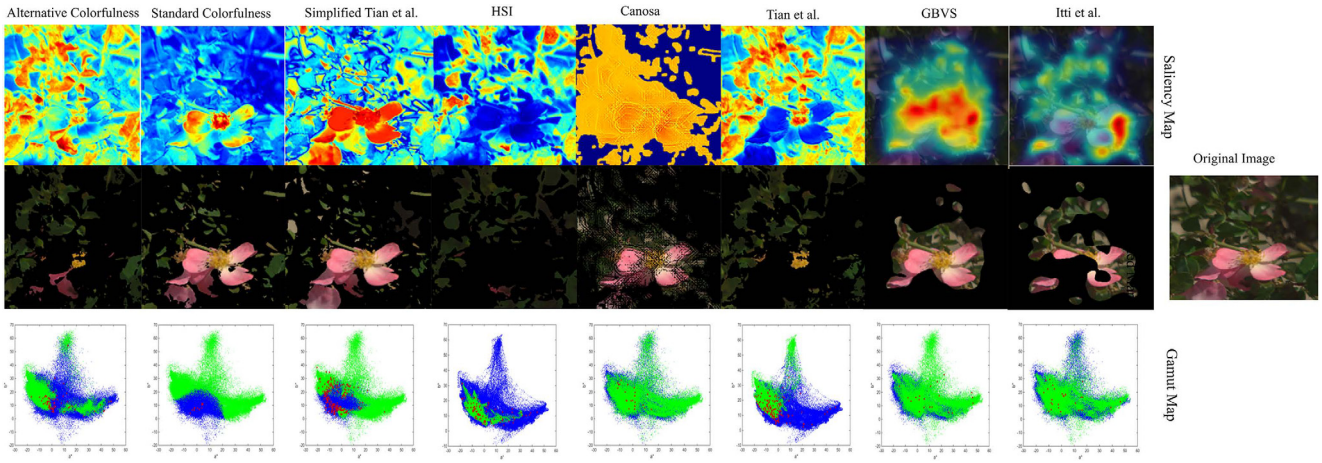

FIGURE 4 Example of the obtained results for (upper row) salient maps, (central row) thresholded salient maps, and (lower row) color gamut expanded in (a^*, b^*) plane

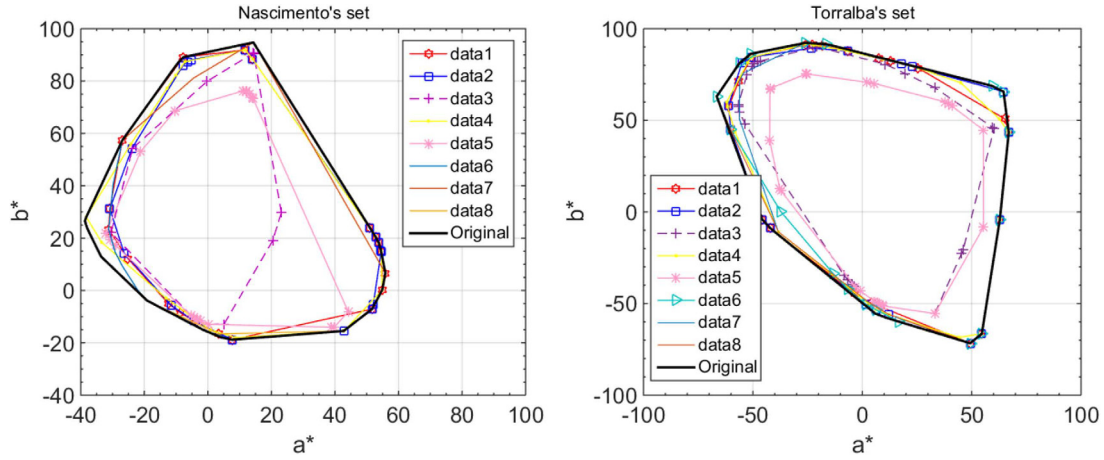

FIGURE 5 Average color gamuts estimated for all salient colors as predicted for each saliency model (data 1–8) in comparison with the original gamut

33.5); on the contrary large values for this distance are obtained for the Modified Tian and HSI models, as expected from the visual inspection of the Figure 4. This could be indicative of some-kind of driven-task (i.e., salient discernible color gamut in our study) conditioning the selection of the best saliency model.

TABLE 1 Hausdorff distances for the 2 image data sets and saliency model tested

Model	Nascimento's set	Torralba's set
Itti model	41.7	57.8
GBVS	54.3	62.3
Tian	44.8	59.5
Canosa	33.5	60.7
HIS	60.0	60.6
Modified Tian	72.0	66.7
Standard colorfulness	55.2	67.8
Alternative colorfulness	42.5	68.3

4.2 | Estimating the number of remarkable salient colors

The Number of Remarkable Salient Colors was estimated by counting the number of centroids in the labeled image $I_{\text{labeled}}(x,y)$. Table 2 shows the average NRSC found for the set of images and across each model described in Section 2.3. Results show how the Itti et al. and GBVS models get a reduced NRSC (number of salient colors around 17–25) with a relatively low standard deviation. On the contrary, the Tian et al. model gets large NRSC with large standard deviation which indicates its strong dependence with the color content of images. The alternative colorfulness feature model clearly fails to obtain a low standard deviation and Tian et al. and Canosa models also get a relatively large standard deviation in comparison with the number NRSC found. By comparing the colorfulness feature definitions in Equations A6 and A7 the presence of L^* at the denominator could explain the more stable values for the standard colorfulness definition in comparison with the alternative formula in Equation A7. Figure 6 shows examples of the spatial locations of each of the salient colors found for each algorithm.

To visualize the remarkable salient colors in each image we used those salient image regions to perform a salient-color-based segmentation using a simple *k*-means clustering.

TABLE 2 Number of remarkable salient colors (NRSC) for the 2 image data sets and saliency model tested

		Itti	GBVS	Tian	Canosa	HSI	Modified Tian	Standard colorfulness	Modified colorfulness
Nascimento	NRSC	25	23	253	30	174	236	142	58
	SD[a]	3	5	125	22	79	112	111	105
Torralba	NRSC	20	17	168	17	134	125	80	31
	SD[a]	4	7	168	12	98	77	76	33

[a]Standard deviation.

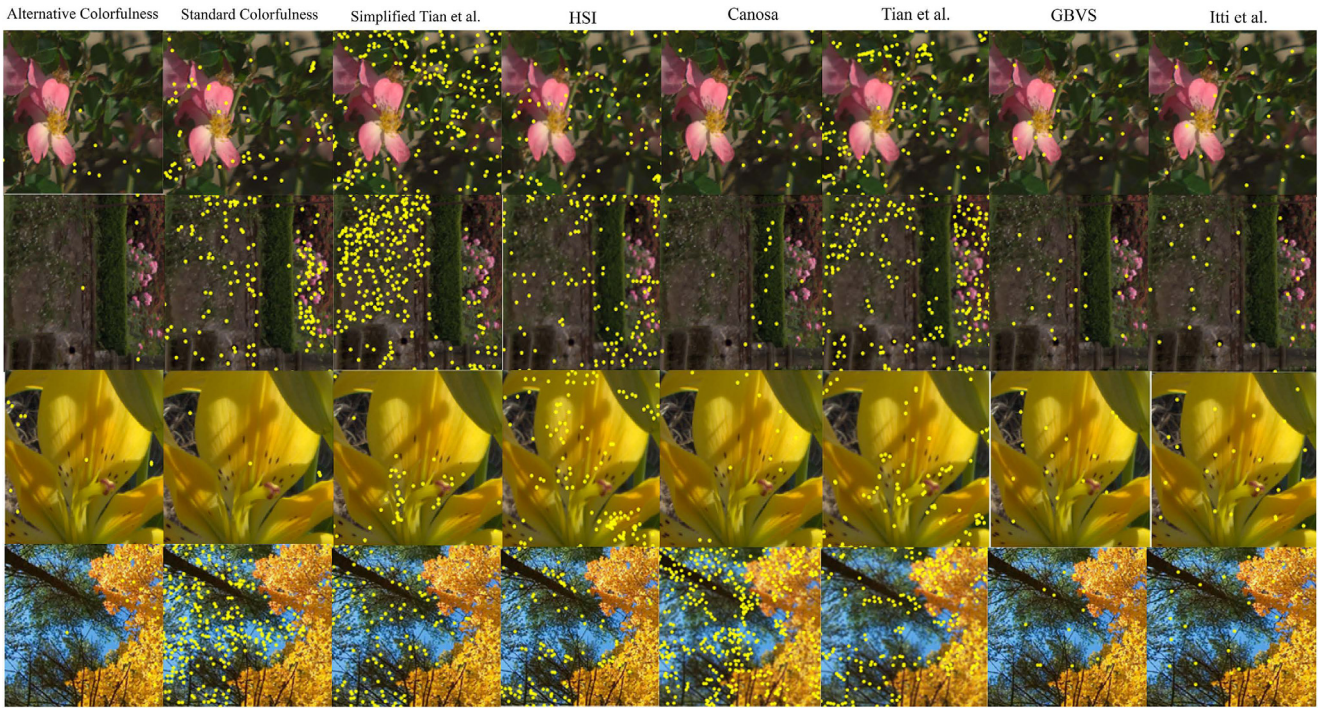

FIGURE 6 Examples of the spatial locations (yellow markers) of each remarkable salient color found for each algorithm; images in the last column are from Torralba's data set

Every salient color, as defined through its (a^*,b^*) chromatic components, was used as a seed for the k-means algorithm, that is, the number of clusters to be partitioned and to quantify how close to regions are to each other. The purpose of this step is to introduce a simple segmentation approach that uses the salient colors to visualize the salient chromatic diversity of color images better. For every salient color the clustering algorithm returns an image labeled by a cluster index as shown in the Figure 7.

4.3 | Applications with artificial scenes

The salient colors computation introduced in this work, which is based on the chromatic salient and discernible chromatic content in images, may be applicable in non-natural scenes such as paintings and/or images including man-made objects. Those kind of objects and scenes were avoided in the previous analysis because scan patterns during those scene viewing can be biased by other oriented tasks (i.e., free viewing vs. ratings of affect and compositional complexity). When an observer looks at a painting, his/her task is usually driven either by the mere observation of the aesthetic of the artistic work or by a close observation of the details of its pictorial style. In both cases there will be a clear saliency map guided by the interest of the painting, characters appearing in it, landscape if any, and so on, and of course the color palette used by the author. It is precisely this latter kind of information that can be highlighted more clearly when we apply our approach because our approach could estimate which colors from the painter's palette will be more remarkable for the observer looking at the painting.

Figures 8 and 9 show examples for an artificial scene and a painting, respectively, when the GBVS model is used prior to detect the salient colors appearing in the images. Results suggest how despite the relative complexity of spatio-

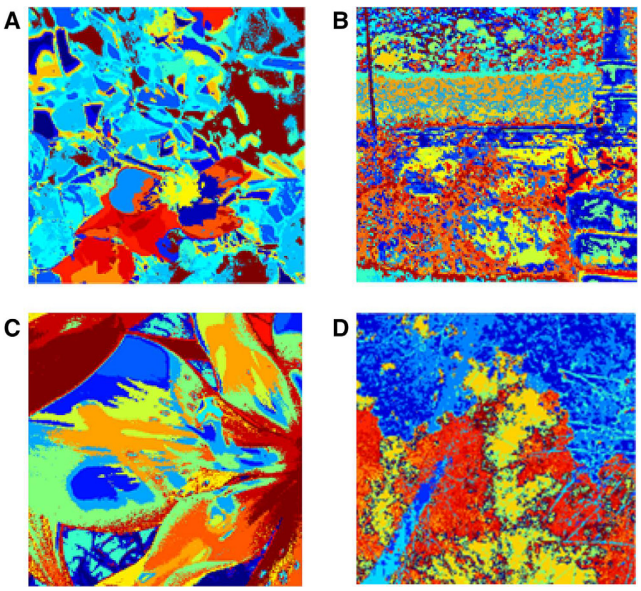

FIGURE 7 Segmentation and visualization examples of the remarkable salient colors using the GBVS model. Labels (A)-(D) correspond to the images shown in rows 1–4 in Figure 6

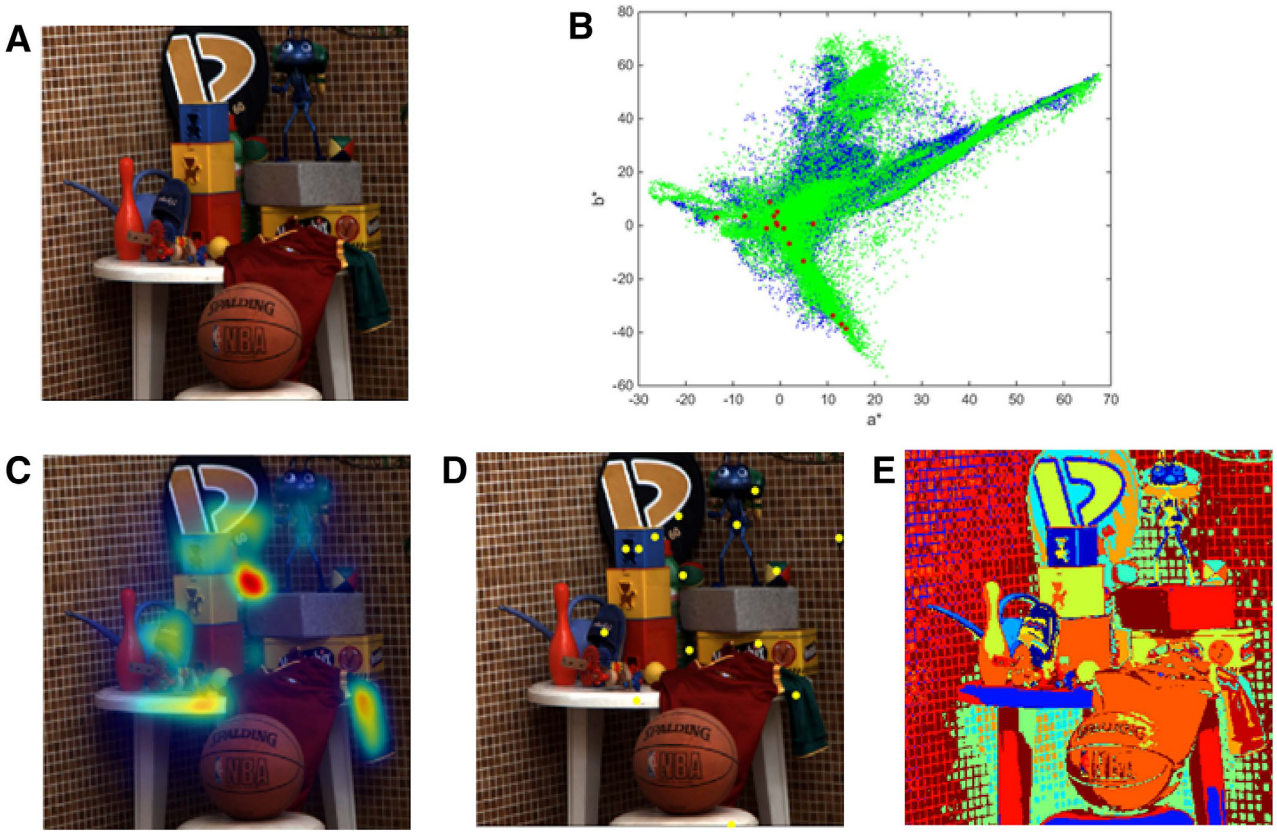

FIGURE 8 (A) Artificial scene, (B) distribution of colors for the global image (blue markers) and salient image regions (green markers) in (a^*,b^*) plane, (C) superimposed saliency map, (D) location of the remarkable salient colors across the image, and (E) labeled image according to the relevant clusters of colors

chromatic content of the scenes the proposed algorithm is able to perform a good image segmentation based on their content, that is, based on their salient colors. In the scene containing toys (Figure 8) salient information is concentrated around the central part of the image, which is an expected result from all visual saliency models, not necessarily around all toys. Nonetheless the 13 salient colors found by the algorithm can segment quite well the color image (Figure 8E). Nevertheless, that segmentation qualitatively contains some errors due to highlights and shadows in the original scene and our proposal was applied without any additional preprocessing of the image (shadow removal, lighting invariants, etc.).

Figure 9 shows an example of the results obtained when a relatively colorful painting containing nature is used. Color gamut of this painting (Figure 9C) is greater than the one expanded by the toy scene (Figure 8C) and now saliency map almost cover the whole scene selecting a total of 8 salient colors. Our labeling algorithm produces the false color image shown in Figure 9E whose segmentation is closer to the original color image than the other example.

Our estimation of salient colors in a scene is qualitatively similar but completely different to other approaches that create a non-photorealistic rendering (NPR) of a photograph. Examples of the NPR technique have been proposed by DeCarlo and Santella[28] who automatically rendered photographs by "stylizing and abstracting photographs that responds in explicit terms to the design goal of clarifying the meaningful visual structure in an image." Using an eye tracking system the authors rendered more details of images at the observers' gaze locations and coarse details in the non-fixated image regions. Those authors look for an abstraction preserving the meaningful of the scene content while we use our model to predict image regions where salient colors appear.

5 | DISCUSSION

We have performed an empirical comparison of 8 state-of-the-art saliency algorithms in computing the chromatic diversity in natural scenes constrained by a saliency-driven task. Despite research on the number of discernible colors,[4,5] so far not much attention has been paid to the influence of visual saliency in the computation of those discernible colors. Instead of processing the massive sensory input, which determines the chromatic diversity of natural scenes, attention guilds us to some "salient"

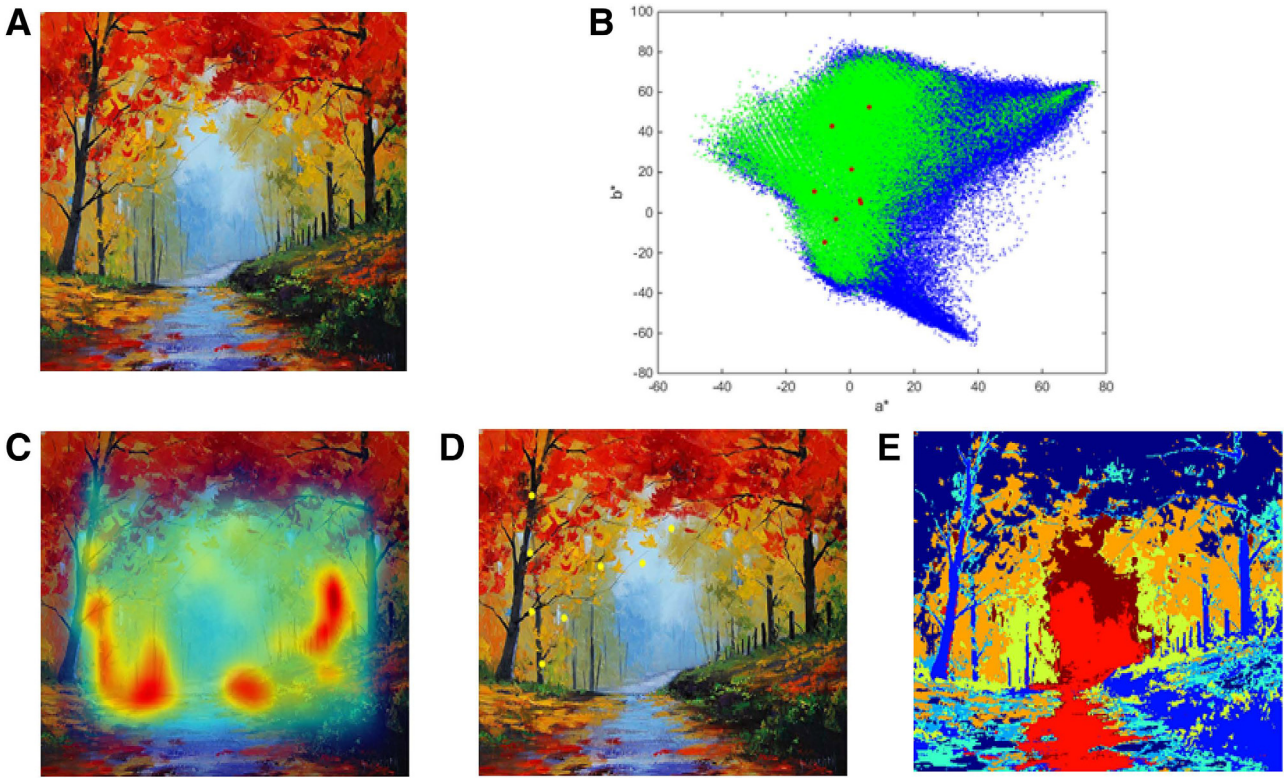

FIGURE 9 (A) Original painting containing nature, (B) distribution of colors for the global image (blue markers) and salient image regions (green markers) in (a^*,b^*) plane, (C) superimposed saliency map, (D) location of the remarkable colors across the image, and (E) labeled image according to the relevant clusters of colors

regions to perform a series of rapid computationally less demanding, localized visual analyses. Attention could be involved in triggering behavior related to recognition and planning and thus reducing the number of salient colors contributing to the total number of discernible colors appearing in an image. Our results suggest that the GBVS saliency model performs best preserving the chromatic diversity content of natural scenes (i.e., low dissimilarity values between the original color gamut and the obtained color gamut after selecting the salient regions). Besides, the influence of local image properties in eye fixations when an observer scans a natural scene is known with around 60% of the variance in observers' detection performance supporting those properties.[29] The starting hypothesis of this work has been precisely that the whole, or part, of that percentage could be in the end diminishing the effective chromatic diversity of natural scenes. Most pattern recognition theories suggest that the visual system must use a specific mechanism to carry out such analysis and estimation, that is, once the basic components of an image are visually processed on the basis of that information the object or visual pattern is identified.[30]

While the mere application of a model of visual saliency has allowed us to calculate the number of salient discernible colors, it is true that this computation is only determined by the local image properties, namely color, contrast and orientation, or equivalently a bottom-up visual processing. On the contrary the estimation of the salient colors appearing in a scene would be the result of a task-driven top-down process. Massaro et al.[31] have shown that gaze is preferably affected by bottom-up processes such as color and complexity when natural scenes are observed. Using an eye tracker system they computed the number of clusters involved in aesthetic judgements of paintings and found around 13 clusters when paintings representing nature were observed. In this respect this number is lower than the number of salient colors we have found in our study. It should be take into account that we used natural images and not nature paintings; nevertheless that number of cluster agrees with the salient chromatic diversity illustrated by the two artificial scenes employed in our study.

The computation of the salient colors has been made under the assumption that salient regions in the image determine such computation. This does not mean that human visual system would estimate those salient colors following a similar procedure, but it is as if we had imposed a specific task to a simulated observer in order to extract salient color targets (or the relevant color "gists" of the images). Results from Castelhano and Henderson[32] suggest that color has a direct connection to the conceptual representation of scenes

by contributing to the activation of scene gist. If salient colors can be considered a scene gist and/or a new color naming paradigm is a matter of further studies.

6 | CONCLUSIONS

We make 3 main contributions in this work. First, a new concept about "salient discernible color" has been introduced to study the chromatic diversity that is really salient in an image. Second, the state-of-the-art of 8 saliency models has been analyzed considering how they process the chromatic diversity in images. Although some very good reviews of visual saliency models can be found in the literature, none of them, up to our knowledge, have included the preservation of the chromatic diversity as one of the metrics to be tested. We have quantified the amount of salient chromatic information, which effectively reduced the number of discernible colors in natural images. And third, we have proposed a novel segmentation and labeling color image annotation that is based on the salient colors appearing in images.

ACKNOWLEDGMENTS

This work was supported by Ministry of Economy and Competitiveness of Spain, under research grant DPI2015–64571-R.

ORCID

Juan Luis Nieves http://orcid.org/0000-0002-3103-8322

REFERENCES

[1] Geisler WS. Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol.* 2008;*59*:167–192.

[2] Barlow HB. Possible principles underlying the transformation of sensory messages. In: Rosenblith WA, eds. *Sensory Communication*. Cambridge, MA: MIT Press; 1961;217–234.

[3] Simoncelli EP, Olshausen BA. Natural images statistics and neural representation. *Annu Rev Neurosci.* 2001;*24*:1193–1215.

[4] Linhares JM, Pinto PD, Nascimento SM. The number of discernible colors in natural scenes. *J Opt Soc Am A* 2008;*25*:2918–2924.

[5] Masaoka K, Berns RS, Fairchild MD, Abed FM. Number of discernible object colors is a conundrum. *J Opt Soc Am A* 2013;*30*:264–277.

[6] Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 2000;*40*(10–12):1489–1506.

[7] Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol.* 1985;*4*:219–227.

[8] Canosa RL. Modeling selective perception of complex natural scenes. *Int J Artif Intell Tools.* 2005;*14*(01n02):233–260. no.

[9] Wu B, Xu L. Integrating bottom-up and top-down visual stimulus for saliency detection in news video. *Multimed Tools Appl.* 2014;*73*:1053–1023.

[10] Privitera MC, Stark LW. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Trans Pattern Anal Mach Intell.* 2000;*22*:970–982.

[11] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell.* 1998;*20*:1254–1259.

[12] Tatler BW, Baddeley RJ, Gilchrist ID. Visual correlates of fixation selection: effects of scale and time. *Vision Res.* 2005;*45*:643–659.

[13] Baddeley R, Tatler B. High frequency edges (but not contrast) predict where we fixate: a Bayesian system identification analysis. *Vision Res.* 2006;*46*:2824–2833.

[14] TianWan MS, Yue L. "A Color Saliency Model for Salient Objects Detection in Natural Scenes", Advances in Multimedia Modeling, vol. *5916*, Lecture Notes in Computer Science, 2010:240–250.

[15] Bylinskii Z, DeGennaro EM, Rajalingham R, Ruda H, Zhang J, Tsotsos JK. Towards the quantitative evaluation of visual attention models. *Vision Res.* 2015;*116*:258–268.

[16] Borji A, Itti L. State-of-the-art in visual attention modeling. *IEEE Trans Pattern Anal Mach Intell.* 2013;*35*:185–207.

[17] Judd T, Durand F, Torralba A. A benchmark of computational models of saliency to predict human fixations. In: MIT Technical Report, USA, 2012.

[18] Frey H-P, König P, Einhäuser W. The role of first- and second-order stimulus features for human overt attention. *Percept Psychophys.* 2007;*69*:153–161.

[19] Frey H-P, Wirz K, Willenbockel V, et al. Beyond correlation: do color features influence attention in rainforest? *Front Hum Neurosci.* 2011;*5*:36–31.

[20] Nascimento SMC, Ferreira F, Foster DH. Statistics of spatial cone-excitation ratios in natural scenes. *J Opt Soc Am A Opt Image Sci Vis.* 2002;*19*(8):1484–1490.

[21] CIE 15:2004 Colorimetry, 3rd Ed., Vienna: CIE Central Bureau, 2004.

[22] Zhao Q, Koch C. Learning saliency-based visual attention: a review. *Signal Process.* 2013;*93*:1401–1407.

[23] Harel J, Koch C, Perona P. Graph-based visual saliency. In: Advances in Neural Information Processing Systems, No. 19 (NIPS. 2006). Cambridge, MA: MIT Press; 2006:545–552.

[24] Parkhurst D, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* 2002;*42*:107–123.

[25] Fairchild M. Color Appearance Models. Hoboken, NJ: John Wiley & Sons. 2013: *87*.

[26] Lübbe E. Colours in the mind—colour systems in reality. In: Taschenbuch, ed. *A Formula for Colour Saturation*. Muster-Schmidt Verlag, Gleichen, Zurich. 2010.

[27] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell.* 1993;*15*:850–863.

[28] DeCarlo D, Santella A. Stylization and abstraction of photographs. *In ACM Trans Graph.* 2002;*21*:769–776.

[29] Amano K, Foster DH. Influence of local scene color on fixation position in visual search. *J Opt Soc Am A* 2014;*31*:A254–A262.

[30] Nobre AC, Kastner S. The Oxford Handbook of Attention. Oxford Library of Psychology. Oxford, UK: Oxford University Press; 2014.

[31] Massaro D, Savazzi F, Di Dio C, et al. When art moves the eyes: a behavioral and eye-tracking study. *PLoS One*. 2012;*7*(5): e37285–e37216.

[32] Castelhano MS, Henderson JM. The influence of color on the perception of scene gist. *J Exp Psychol Hum Percept Perform*. 2008;*34*:660–675.

AUTHOR BIOGRAPHIES

JUAN LUIS NIEVES received M.S. and Ph.D. degrees in physics from the University of Granada, Granada, Spain, in 1991 and 1996, respectively. He is currently Full Professor with the Department of Optics, Science Faculty, at University of Granada, where he conducts research in the Color Imaging Laboratory. His current research interests include computational color vision (color constancy, human visual system processing of spatio-chromatic information), and spectral analysis of color images. Dr. Nieves is the President of the Spanish Color Committee, and representative of this Committee in the International Color Association (AIC), and is currently the coordinator of the Erasmus+ Joint Master Degree "Color in Science and Industry (COSI)."

JAVIER ROMERO is currently Full Professor of Optics and Photonics with the University of Granada, where he is the Head of the Color Imaging Group. His current research interests include colorimetry, color vision, and multispectral and color imaging. Dr. Romero is the Past-President of the International Color Association (AIC).

How to cite this article: Nieves JL, Romero J. Heuristic analysis influence of saliency in the color diversity of natural images. *Color Res Appl.* 2018;43:713–725. https://doi.org/10.1002/col.22235

APPENDIX

The saliency map is a biologically plausible model for bottom-up overt attention. In the review of the model by Itti et al.[6] the visual saliency maps are topographical codifications of fixation position in visual search over the entire scene based on different image features such as luminance, orientation or color. Models 1, 2, and 4 in Section 3.2 are based on the flowchart illustrated in the Figure 10.

The Model 3 is based on the computation of different contrast-based features[14]; its 5 M_i saliency maps can be computed as:

- For contrast-based features:

$$M_i(x,y) = 1 - \exp\left\{ - \frac{|F_i(x,y) - \bar{F}_i(x,y)|}{|F_i(x,y) - \bar{F}_i(x,y)|} \right\} \quad \text{(A1)}$$

where i stands for the hue, saturation, and intensity channels of the image.

- For dominance-based features:

$$M_4(x,y) = \begin{cases} I(x,y)S(x,y)\cos(\theta)\delta & \text{if } \frac{\sqrt{2}}{2} \le \cos(\theta) \le 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{(A2)}$$

$$M_5(x,y) = I(x,y)S(x,y)\delta \quad \text{(A3)}$$

where $I(x, y)$ and $S(x, y)$ denote the intensity and the saturation image channels, respectively, θ is the hue for pixel (x, y) and normalized to $[0, 2\pi]$, and δ is a constant value.

In the Model 5, the saliency map computation is based on an equal weight of the hue, saturation, and intensity components of each image defined as:

$$M(x,y) = \sqrt{H(x,y)^2 + S(x,y)^2 + I(x,y)^2} \quad \text{(A4)}$$

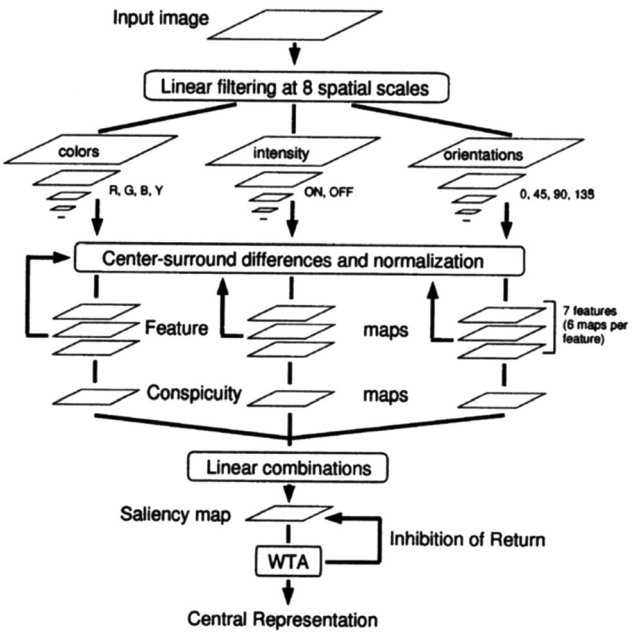

FIGURE 10 Scheme of the original model of saliency-based visual attention, adapted from Itti and Koch[6]

where $H(x,y)$, $S(x,y)$, and $I(x,y)$ are the hue, saturation, and intensity values at pixel (x,y) of the image.

The simplified Model 6 uses only the contrast-based features related to the contrast of hue, contrast of saturation, and contrast of intensity as defined by Tian et al.[14]:

$$M(x,y)=\sqrt{M_H(x,y)^2+M_S(x,y)^2+M_I(x,y)^2} \tag{A5}$$

where $M_H(x,y)$, $M_S(x,y)$, and $M_I(x,y)$ are the saliency maps for the contrast-based features for hue, saturation, and intensity, respectively, as defined by Equation A1.

The Model 7 computes the colorfulness as follows:

$$M(x,y)=\frac{\sqrt{a^*(x,y)^2+b^*(x,y)^2}}{L^*(x,y)} \tag{A6}$$

Finally, the Model 8 uses the following formula to compute an alternative colorfulness attribute to compute the salient regions:

$$M(x,y)=\frac{\sqrt{a^*(x,y)^2+b^*(x,y)^2}}{\mathrm{std}\left[a^*(x,y)^2+b^*(x,y)^2\right]} \tag{A7}$$

where std stands for the standard deviation of the quantity within the brackets.