

# Contribution of color in saliency model for videos

Shahrbanoo Hamel<sup>1</sup> · Nathalie Guyader<sup>1</sup> · Denis Pellerin<sup>1</sup> · Dominique Houzet<sup>1</sup>

Received: 31 October 2014 / Revised: 2 March 2015 / Accepted: 9 March 2015 / Published online: 24 March 2015  
© Springer-Verlag London 2015

**Abstract** Much research has been concerned with the contribution of the low-level features of a visual scene to the deployment of visual attention. Bottom-up saliency models have been developed to predict the location of gaze according to these features. So far, color besides intensity, contrast and motion is considered as one of the primary features in computing bottom-up saliency. However, its contribution in guiding eye movements when viewing natural scenes has been debated. We investigated the contribution of color information in a bottom-up visual saliency model. The model efficiency was tested using the experimental data obtained on 45 observers who were eye-tracked while freely exploring a large dataset of color and grayscale videos. The two datasets of recorded eye positions, for grayscale and color videos, were compared with a luminance-based saliency model (Marat et al. *Int J Comput Vis* 82:231–243, 2009). We incorporated chrominance information to the model. Results show that color information improves the performance of the saliency model in predicting eye positions.

**Keywords** Color information · Visual saliency · Video · Eye tracking

## 1 Introduction

The mechanism of visual attention allows selecting the relevant parts of a visual scene at the very beginning of exploration. The selection is driven by the properties of the visual stimulus through bottom-up processes, as well as by the goal of observer through top-down processes [6, 15]. Visual attention models tend to predict the parts of the scene that are likely to deploy the attention [11, 14, 20, 21]. Most of the models are bottom-up models based on the feature integration and guided search theories [30, 32]. These theories stipulate that some elementary salient visual features such as intensity, color, depth and motion are processed in parallel at a pre-attentive stage, subsequently combined to drive the focus of attention. This approach is in accordance with the physiology of the visual system. Hence, in almost all the models of visual attention, low-level features such as intensity, color and spatial frequency are considered to determine the visual saliency of regions in static images, whereas motion and flicker are also considered in the case of dynamic scenes [14, 20, 21]. More recently, the contribution of different features like color in guiding eye movements when viewing natural scenes has been debated. Some studies suggested that color has little effect on fixation locations [2, 10, 13], which brings to question the necessity of the inclusion of color features in the saliency models [8]. In this study, we investigated the contribution of color information in predictive power of saliency model by incorporating color to a luminance-based model of saliency [21]. We also identified and compared the salient regions of a dataset of color videos and same videos in grayscale, through an eye-tracking experiment.

---

This research was supported by the Rhone-Alpes region (France). We thank A. Rahman for the GPU implementation of saliency model of Marat et al. We also thank D. Alleysson and D. Meary for providing us with spectrometer measurements.

---

✉ Shahrbanoo Hamel  
Shahrbanoo.Hamel@gipsa-lab.grenoble-inp.fr  
Nathalie Guyader  
Nathalie.Guyader@gipsa-lab.grenoble-inp.fr  
Dominique Houzet  
Dominique.Houzet@gipsa-lab.grenoble-inp.fr

<sup>1</sup> GIPSA-lab, UMR 5216, Grenoble, France

## 2 Method

### 2.1 Saliency model

The saliency model of Marat et al. [21] is a biologically plausible model that imitates the human visual system from retina to cortex. The model is only based on the luminance visual information. As shown in Fig. 1, in a preprocessing step, the luminance visual information is elaborated by retina-like filters and then is decomposed using cortical-like filters. The luminance model is consisted of two pathways: one processes the luminance-static information that provides luminance-static saliency map ( $M_{ls}$ ), and one processes the luminance motion information that provides luminance-dynamic saliency map ( $M_{ld}$ ). For luminance-static and luminance-dynamic processing steps, the input grayscale image is obtained from Eq. 9 that is detailed in Sect. 2.2. We incorporated the color information to the model by adding a chrominance-static pathway that provides the chrominance-static saliency map ( $M_{cs}$ ).

The input image is decomposed to a luminance and two chromatic opponent images. There are several color spaces proposing different combination of cone responses to define the principal components of luminance and opponent colors, red–green (RG) as well as blue–yellow (BY) [31]. The color space proposed by Krauskopf et al. [18] is one of

the validated representations to encode visual information where the orthogonal directions, A, Cr1 and Cr2, represent luminance, chromatic opponent red–green and chromatic opponent yellow–blue, respectively. Equation 1 is used to compute A, Cr1 and Cr2. In our model, we used Cr1 and Cr2 to compute the chrominance-static saliency map.

$$\begin{pmatrix} A \\ Cr1 \\ Cr2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -0.5 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix} \quad (1)$$

where L, M and S correspond to the response of the three types of cones of the human eye; their name was chosen because of their sensitivity at long, medium and short wavelengths of the light. L, M and S values are calculated from tristimulus values of 1931 CIE XYZ color space as follows:

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.4002 & 0.7076 & -0.0808 \\ -0.2263 & 1.1653 & 0.0457 \\ 0 & 0 & 0.9182 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (2)$$

The different steps of the saliency model that follow the input image decomposition are presented below.

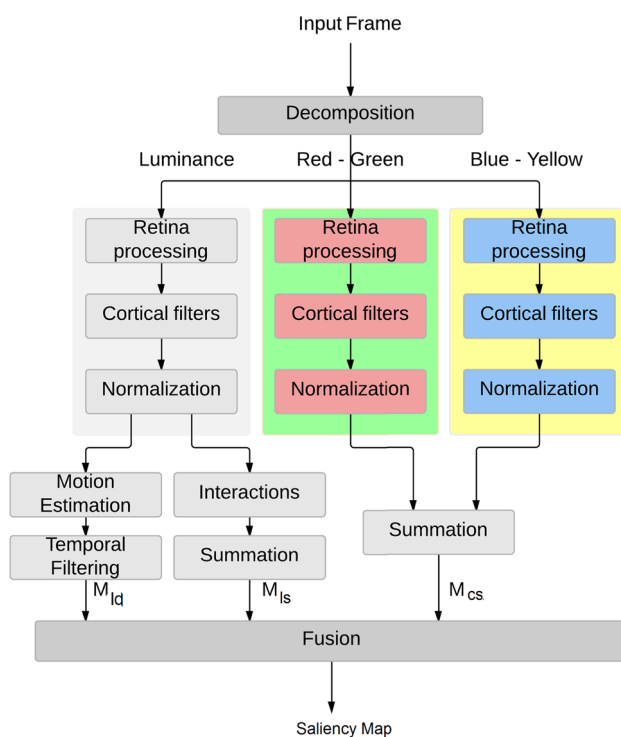
#### 2.1.1 Retina-like filters

The retina, which has been described in detail in [21], roughly simulates the functioning of retinal cells without taking into account the spatially variant resolution of the retina photoreceptors. The retina-like filters decompose the input signals into two main outputs: a parvocellular-like output that enforces the high spatial frequencies to enhance the contrasts, and a magnocellular-like output that conveys lower spatial frequencies. The first output is used to compute the luminance-static saliency maps and the latter to compute the luminance-dynamic saliency maps.

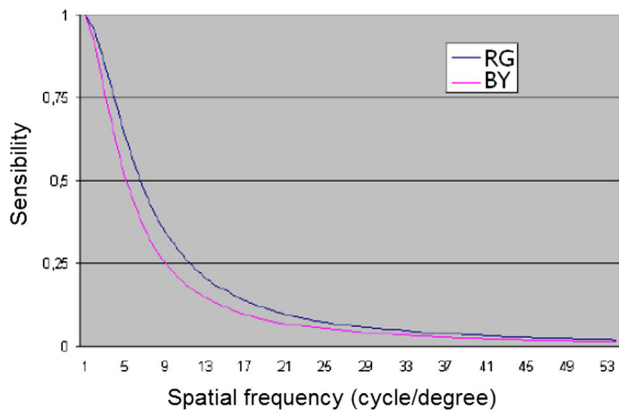
As for the chrominance information, the retina is modeled by low-pass filters that the transformation functions reproduce the contrast sensitivity functions of retina for red–green and blue–yellow opponents as shown in Fig. 2, and also a Gaussian low-pass filter.

#### 2.1.2 Cortical-like filters

The frequency and orientation selectivity of visual cortex are modeled by a bank of Gabor filters. Gabor filters are oriented band-pass filters that are characterized by their frequency selectivity and orientation. Each Gabor filter  $G_{ij}$  (Eq. 3) is defined by its central radial frequency  $f_j$  and its standard deviations  $\sigma_{ij}^\theta$  and  $\sigma_{ij}^f$  in orientation  $\theta_j$  and its orthogonal orientation, respectively,  $i = 1, \dots, N_\theta$ ,  $j = 1, \dots, N_f$ ,  $\frac{f_j}{f_{j-1}} = 2$  and  $f_{N_f} = 0.25$ .



**Fig. 1** The spatiotemporal saliency model.  $M_{ld}$  is luminance-dynamic map, and  $M_{ls}$  and  $M_{cs}$  are luminance-static and chrominance-static maps, respectively



**Fig. 2** The normalized contrast sensibility functions of the color components Cr1 and Cr2, from [19]

$$G_{ij}(u, v) = \exp \left\{ - \left( \frac{(u' - f_j)^2}{2(\sigma_{ij}^u)^2} + \frac{v'^2}{2(\sigma_{ij}^v)^2} \right) \right\} \quad (3)$$

where,  $u' = u \cos\theta + v \sin\theta$  and  $v' = v \cos\theta - u \sin\theta$ .

For luminance information, the initial model, proposed by Marat and colleagues [21], uses six orientations and four frequencies. Hence,  $N_\theta = 6$  and  $N_f = 4$ . Since the amplitude spectra of the two color-opponent Cr1 and Cr2 images do not have as many specific orientations as the amplitude spectra of the luminance image [3], for both chrominance information, Cr1 and Cr2, only Gabor filters with four orientations are used ( $0^\circ, 45^\circ, 90^\circ$  and  $135^\circ$ ). Because human visual system is less sensitive to the high spatial frequencies of chrominance information [12], only two lowest frequencies are chosen (0.25 and 0.125 cycle per degree).

### 2.1.3 Static saliency maps

Two operations are carried out to create one luminance-static saliency map from the output of cortical-like filters,  $M_{u,v}$  intermediate maps: the interactions and the normalization. The interactions between neighboring pixels of the intermediate maps models the lateral neural connections of visual cortex. They are modeled as linear combination of neighboring pixels. The interactions, depending on the orientation or the frequency, may be excitatory when in the same direction, or inhibitory otherwise.

$$M_{u,v} = M_{u,v} \cdot w \quad (4)$$

where,

$$w = \begin{bmatrix} 0 & -0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0.0 & -0.5 & 0 \end{bmatrix} \quad (5)$$

Then the intermediate maps are normalized using the method proposed by Itti et al. [14]. First, each intermediate map is normalized to [0 1], then it is multiplied by  $(\max(M_{u,v}) - \overline{M_{u,v}})^2$ . Then all values lower than 0.2 are set to zero. The normalization enforces the saliency of the regions that are different from their surrounding, by unifying the dynamic range of the intermediate maps. Then a luminance-static saliency map,  $M_{ls}$ , is obtained by summing up all the normalized maps.

To compute the chrominance-static saliency map, first the red–green and blue–yellow intermediate maps are normalized to [0 1] and then are summed up to obtain a chrominance saliency  $M_{cs}$ .

### 2.1.4 Dynamic saliency map

The dynamic saliency is related to the moving objects of the scene. The magnocellular output is used to detect the objects that are moving against the background. A differential approach is used for motion estimation by solving a system of optical flow equations [5]. For every frame, a motion vector is defined per pixel. Only the modulus of the vector is used to define the dynamic saliency of a region, assuming that the motion saliency map of a region is proportional to its speed against the background. Then a temporal median filter is applied to five successive frames to remove the possible noisy detected motions. The output of temporal filtering is considered as luminance-dynamic saliency map,  $M_{ld}$  (Fig. 1).

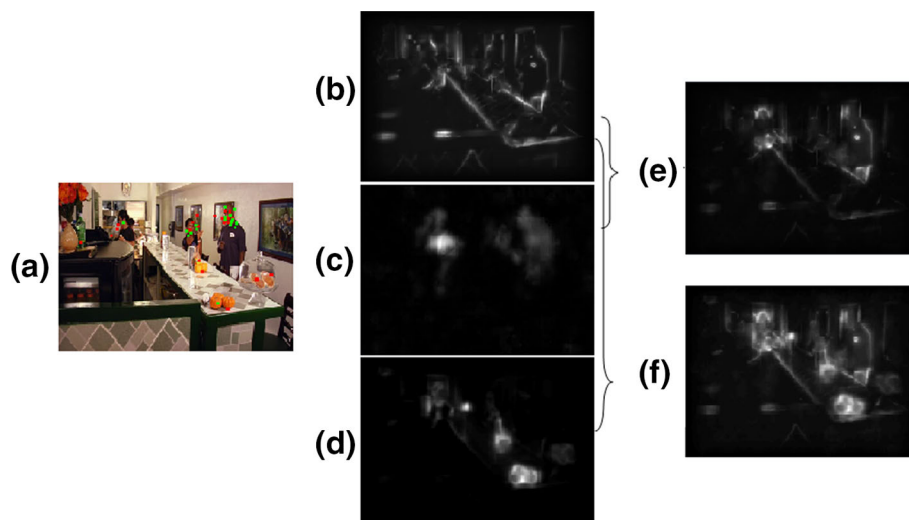
Chrominance-static saliency map  $M_{cs}$ , luminance-static saliency map  $M_{ls}$  and luminance-dynamic saliency map,  $M_{ld}$ , after normalizing to [0 1], are combined, according to Eq. 6, to obtain a master spatiotemporal saliency map per video frame. This map predicts the salient regions i.e., the regions that stand out in a visual scene.

$$\text{Saliency map} = \alpha M_{ls} + \beta M_{ld} + M_{cs} + \alpha\beta(M_{ls} \cdot M_{ld}) \quad (6)$$

where  $\alpha$  and  $\beta$  are the max of  $M_{ls}$  and skewness of  $M_{ld}$ , respectively, and  $M_{ls} \cdot M_{ld}$  is a pixel to pixel multiplication.

The weights of maps in Eq. 6 were found to result a good fusion regarding the fact that the saliency maps from both the static and dynamic pathways exhibit different characteristics, i.e., static saliency map has larger salient regions based on textures, whereas dynamic saliency map has smaller salient regions depending on the moving objects [21, 24]. Static and dynamic maps are modulated using maximum and skewness, respectively. The reinforcement parameter  $\alpha\beta$  is used to include the regions that have low motion, but include large salient regions in static saliency map. Figure 3 shows an example frame and its intermediate and final saliency maps.

**Fig. 3** Saliency maps: **a** An example frame, **b** luminance-based static map  $M_{ls}$ , **c** luminance-based dynamic map  $M_{ld}$ , **d** chrominance-based static map  $M_{cs}$ , **e** fusion of  $M_{ls}$  and  $M_{ld}$ , **f** fusion of  $M_{ls}$ ,  $M_{cs}$  and  $M_{ld}$



The performance of the model is also compared with one of the reference saliency models, Itti and Koch saliency model [14, 17].

### 2.1.5 GPU implementation

The saliency model presented above with static (luminance-based), dynamic (luminance-based) and chrominance pathways is compute-intensive. Rahman et al. [25] have proposed a parallel adaptation of luminance-based pathways onto GPU. They applied several optimizations subtending to a real-time solution on multi-GPU. We included the parallel adaptation of chrominance pathway to this GPU implementation maintaining the real-time solution.

The *NVIDIA CUDA fast Fourier transform* library (*cuFFT*) was used to perform the complex Fourier transformations. The reductions were carried out using *Thrust* library, an interface to many GPU algorithms and data structures. Such as the implementation of luminance-static and luminance-dynamic pathways, chrominance pathway was tested on a 2.67 GHz quad-core system with 10 GB of main memory and Windows 7 running on it. CUDA v3.0 programming environment on NVIDIA Geforce GTX 480 was used.

### 2.1.6 NSS metric

A common metric to compare experimental data to computational saliency maps is the normalized scanpath saliency (NSS) [15]. We used this metric to compare C and GS eye positions to their equivalent saliency maps. To compute this, first the saliency maps were normalized to zero mean and unit standard deviation. The NSS value of frame  $k$  corresponds to averaged saliency values at the locations of eye positions on the normalized saliency map  $M$  as shown in Eq. 7:

$$NSS(k) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma_k} (M(X_i) - \mu_k) \quad (7)$$

where  $N$  is the number of the eye positions,  $M(X_i)$  is the saliency value of the eye position ( $X_i$ ), and  $\mu_k$  and  $\sigma_k$  are the mean and standard deviation of the initial saliency map of frame  $k$ . A high positive value of NSS indicates that the eye positions are located on the salient regions of the computational saliency map. A NSS value close to zero represents no relation between eye position and the computational saliency map, while a high negative value of NSS means that eye positions were not located on the salient regions of computational saliency map.

## 2.2 Eye-tracking experiment

This research focused on the contribution of color information in human visual attention. We studied from one side the contribution of color information in a computational model of attention, and from the other side, we analyzed its influence on the eye movements. To study the influence of color on the visual attention, we carried out an eye-tracking experiments. We collected and compared the eye movements data of people who observed the video stimuli in two conditions: color and grayscale.

An EyeLink 1000 from SR research was used to record the eye positions in a pupil tracking mode. The stimuli consisted of 65 short video extracts of 3–5 sec, called video snippets. Video snippets were extracted from various open source color videos. The stimuli measured  $640 \times 480$  pixels, subtending a visual angle of  $25^\circ \times 19^\circ$  at a fixed viewing distance of 57 cm. The temporal resolution of video snippets was 25 frames per sec. The chosen snippets contain two different types of stimuli: person-present scenes (45 snippets) and person-absent

scenes (20 snippets). Person-present scenes include video snippets containing one, two or more persons. The stimuli were observed by 45 volunteers (25 women and 20 men, aged from 25 to 39 years,  $M = 26$ ,  $SD = 4$ , 9) that took part in the experiment. All reported normal or corrected to normal visual acuity, while their normal color vision was tested using the Ishihara test on the experimental display. Each experiment session consisted of two parts. During the first part, the participants watched one half of the video clips in one stimulus condition (color/grayscale), and during the second part, the participants watched the other half of videos in the other condition (grayscale/color).

The two conditions of the video stimuli used in this experiment had to be controlled to achieve the reliable results. In fact, the grayscale version of the video stimuli must preserve the most the features of the original color video. But, color to grayscale conversion is a lossy operation that modifies the luminosity features of the video stimuli. The goal of different grayscale conversion methods is to save the most possible information from the original image.

According to [4], the grayscale conversions could be divided in two main categories: *functional* and *optimizing*. *Functional* methods are pixel-wise methods that process an image locally and compute, for each pixel, a grayscale value from the chromatic values using a given function. The optimizing methods are more advanced models that consider the whole image properties and global characteristics to compute the grayscale image that preserve the most the features of the original image.

We used a simple functional method based on the weighted sum of color channels (Eq. 8). To ensure the luminance matching between color and grayscale stimuli, the right side of relation 8 was fitted to the standard observer luminosity function,  $V(\lambda)$ .

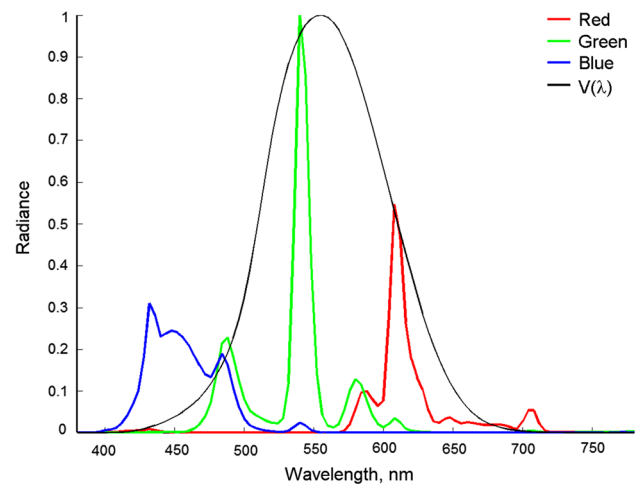
$$Y = w_r \times R + w_g \times G + w_b \times B \quad (8)$$

where  $w_r$ ,  $w_g$  and  $w_b$  were computed according to the radiance of color channels, as the weighted sum in Eq. 8 be equal to the luminosity function of the standard observe.

Figure 4 shows the relative power/radiance of each channel maximum output measured for the LCD display as well as the luminosity function,  $V(\lambda)$ , of standard observer. However, the  $V(\lambda)$  corresponds to the average standard observer, while the response of photocells varies from one observer to another and the random cone mosaic of human eye might affect equiluminance thresholds [1].

A matrix operation is performed to compute the weights. Solving the equation for the measured data, we obtained the following grayscale conversion (Eq. 9):

$$Y = 0.5010 \times R + 0.4911 \times G + 0.0079 \times B \quad (9)$$



**Fig. 4** The relative power/radiance of each channel maximum output measured for the LCD display as well as the luminosity function,  $V(\lambda)$ , of standard observer

In Eq. 9, the weights are normalized to sum to one. Therefore, when  $R$ ,  $G$  and  $B$  values are equal to one, the  $Y$  value is equal to one as well.

### 2.2.1 Eye position analysis metrics

**Dispersion** To evaluate variability of eye positions between observers, we used a metric called *dispersion* [21, 28]. Dispersion was calculated separately for each frame for color eye positions ( $C$  positions) ( $D_C$ ) and grayscale eye positions ( $GS$  positions) ( $D_{GS}$ ). Lower values of dispersion correspond to subjects' eye positions located close to one another, interpreted as high inter-subject consistency.

**Clustering** Salient regions of a visual scene can be identified as the locations fixated by a group of subjects at the same moment of observation. These regions can be estimated by clustering the eye positions of different subjects on each frame [7, 9, 29]. Here, we clustered the eye positions to compare the experimental salient regions in color and grayscale conditions using mean-shift clustering method [29]. This method requires a distance parameter to be adjusted. Because the size of video clips was constant, we empirically set this distance to 75 pixels, equal to nearly 3 degrees of visual angle.

## 3 Results

### 3.1 Saliency model

First, we studied whether *luminance-based saliency model* [21] predicts the eye positions in both conditions with equal efficiency. Then we performed  $NSS$  analysis, but using the model of saliency with chrominance. As shown in Table 1, color information improves significantly the performance of



**Table 1** NSS results for Marat et al. model and Itti and Koch saliency model with and without color features

	Marat		Itti	
	Luminance	Luminance+chrominance	Luminance	Luminance+chrominance
NSS				
C positions	0.59	1.18	0.91	0.95
GS positions	0.60	1.17	0.93	0.97

**Table 2** Timings of sequential (C and Matlab) and parallel (GPU) implementations in ms

	$M_{sl}$	$M_{cl}$	$M_{dl}$
MATLAB	34.01	22.67	237.03
C	10.73	7.15	31.24
CUDA	0.04	0.03	0.12

presented model for both C and GS positions ( $GS : t(63) = 4.5, p < 0.01, C : t(63) = 4.86, p < 0.01$ ), while it improves slightly the performance of the model of Itti and Koch [14].

In addition, as presented in Table 2, GPU implementation of chrominance-static pathway, similar to luminance-static pathway, results in a significant speedup over Matlab and C implementations.

### 3.2 Eye positions

The dispersion of color eye positions is significantly higher than grayscale (5.1 vs. 4.8,  $t(63) = 2, 5804, p < 0.01$ ). This

raw result shows that there is more variability between the eye positions of observers when viewing color videos. Yet, a large dispersion might be observed in two different situations: (1) when all observers look at different areas or (2) when there are several distant clusters of eye positions. The mean number of clusters on color snippets was significantly higher than grayscale (5.1 vs. 4.8,  $t(63) = 2.6, p < 0.01$ ). The result indicates that the high dispersion value of C positions is not due to the high variability of the eye positions, but related to the higher number of regions of interest in color stimuli. However, main clusters were superimposed between C and GS positions. Figure 5 shows the subject regions of interest on an example frame identified by clustering the C positions and GS positions.

### 3.3 Conclusion

In the present manuscript, we have used eye-tracking data; these data allow us to validate the proposed saliency model and more specifically to quantify the contribution of color in the saliency model to predict eye fixations. During the

**Fig. 5** Example of the regions of interest identified by clustering the eye positions. From left to right, first row an example frame in color and grayscale. Second row the corresponding regions of interest of C positions and GS positions (color figure online)



experiment, observers were asked to freely explore video clips in color and in grayscale conditions.

Using a clustering method, we identified the regions of interest that are fixated the most by observers. We found that faces and moving objects correspond to very attractive regions. This result was already described in previous papers [22, 26, 27] for faces and [16, 21, 23] for moving objects. We obtained similar results for both color and grayscale eye positions. However, we found more regions of interest for color stimuli. Due to these results, we have integrated color information into a bio-inspired saliency model proposed by Marat et al. [21].

Results show that indeed color information improves significantly the performance of the model in predicting eye positions for both grayscale and color stimuli, while a better prediction power was expected for color stimuli. This might be due to the fact that the major regions of interest are common in both stimulus conditions, but are better enhanced when employing color information. Yet, the incorporation of color information into the model is not optimized. Because the regions of interest are not always located on colored zones, but their neighboring [20]. Whether reinforcement of luminance saliency according to the color information of neighboring zones can improve the predictive power of saliency model remains to be determined.

## References

- Alleysson, D., Meary, D.: Neurogeometry of color vision. *J. Physiol. Paris* **106**, 284–296 (2012)
- Baddeley, R.J., Tatler, B.W.: High frequency edges (but not contrast) predict where we fixate: a Bayesian system identification analysis. *Vis. Res.* **46**(18), 2824–2833 (2006)
- Beaudot, W.H.A., Mullen, K.T.: Orientation selectivity in luminance and color vision assessed using 2-d bandpass filtered spatial noise. *Vis. Res.* **45**(6), 687–696 (2005)
- Benedetti, L., Corsini, M., Cignoni, P., Callieri, M., Scopigno, R.: Color to gray conversions in the context of stereo matching algorithms. *Mach. Vis. Appl.* **57**(2), 254–348 (2010)
- Bruno, E., Pellerin, D.: Robust motion estimation using spatial Gabor-like filters. *Signal Process.* **82**, 297–309 (2002)
- Connor, C.E., Egeth, H.E., Yantis, S.: Visual attention: bottom-up versus top-down. *Curr. Biol.* **14**, 850–852 (2004)
- Coutrot, A., Guyader, N., Ionescu, G., Caplier, A.: Influence of soundtrack on eye movements during video exploration. *J. Eye Mov. Res.* **5**(4), 1–10 (2012)
- Dorr, M., Martinetz, T., Barth, E.: Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10**(10), 1–17 (2010)
- Follet, B., Le Meur, O., Baccino, T.: New insights on ambient and focal visual fixations using an automatic classification algorithm. *iPerception* **2**(6), 592–610 (2011)
- Frey, H.P., Honey, C., König, P.: What's color got to do with it? the influence of color on visual attention in different categories. *J. Vis.* **11**(3), 1–15 (2008)
- Frintrop, S.: VOCUS: A visual attention system for object detection and goal-directed search. Ph.D. thesis, Rheinische Friedrich-Wilhelms-Universität Für Informatik und Fraunhofer Institut Für Autonome Intelligente Systeme (2006)
- Gegenfurtner, K.R.: Cortical mechanisms of colour vision. *Nat. Rev. Neurosci.* **4**(7), 563–572 (2003)
- Ho-Phuoc, T., Guyader, N., Guérin-Dugué, A.: When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *J. Vis.* **12**(2), 1–13 (2012)
- Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998)
- Itti, L.: Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Vis. Cognit.* **12**, 1093–1123 (2005)
- Itti, L., Baldi, P.: Bayesian surprise attracts human attention. *Vis. Res.* **49**(10), 1295–1306 (2009)
- Klab: <http://www.klab.caltech.edu/harel/share/gbvs.php>
- Krauskopf, J., Williams, D.R., Heeley, D.W.: Cardinal direction of color space. *Vis. Res.* **22**, 1123–1131 (1982)
- Le Callet, P.: Critères objectifs avec référence de qualité visuelle des images couleur. Ph.D. thesis, Université de Nantes (2001)
- Le Meur, O., Le Callet, P., Barba, D.: Predicting visual fixations on video based on low-level visual features. *Vis. Res.* **47**(19), 2483–2498 (2007)
- Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., Guérin-Dugué, A.: Modelling spatio-temporal saliency to predict gaze direction for short videos. *Int. J. Comput. Vis.* **82**(3), 231–243 (2009)
- Marat, S., Rahman, A., Pellerin, D., Guyader, N., Houzet, D.: Improving visual saliency by adding 'face feature map' and 'center bias'. *Cognit. Comput.* **5**(1), 63–75 (2013)
- Mital, P.K., Smith, T.J., Hill, R.L., Henderson, J.M.: Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognit. Comput.* **3**(1), 5–24 (2011)
- Rahman, A.: Face perception in videos: contributions to a visual saliency model and its implementation on GPUs. Ph.D. thesis, Univ. Grenoble Alpes, France (2013)
- Rahman, A., Houzet, D., Pellerin, D., Marat, S., Guyader, N.: Parallel implementation of a spatio-temporal visual saliency model. *Real-Time Image Process.* **6**(1), 3–14 (2010)
- Rahman, A., Pellerin, D., Houzet, D.: Influence of number, location and size of faces on gaze in video. *J. Eye Mov. Res.* **7**(2), 1–11 (2014)
- Rousselet, G.A., Macé, M.J.M., Fabre-Thorpe, M.: Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J. Vis.* **3**(6), 440–455 (2003)
- Salvucci, D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. *Proc. Symp. Eye Track. Res. Appl.* **469**(1), 71–78 (2000)
- Santella, A., DeCarlo, D.: Robust clustering of eye movement recordings for quantification of visual interest. In: eye tracking research and applications (ETRA) symposium (2004)
- Treisman, A.M., Gelade, G.: A feature integration theory of attention. *Cognit. Psychol.* **12**, 97–136 (1980)
- Trémeau, A., Fernandez-Maloigne, C., Bonton, P.: Image numérique couleur. De l'acquisition au traitement, chap. 2, pp. 32–43. Dunod, Paris (2004)
- Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 419–433 (1989)