

Does color influence eye movements while exploring videos?

Shahrbanoo Hamel

Gipsa-lab, CNRS

Grenoble-Alpes University, France

Denis Pellerin

Gipsa-lab, CNRS

Grenoble-Alpes University, France

Dominique Houzet

Gipsa-lab, CNRS

Grenoble-Alpes University, France

Nathalie Guyader

Gipsa-lab, CNRS

Grenoble-Alpes University, France

Although visual attention studies consider color as one of the most important features in guiding visual attention, few studies have investigated how color influences eye movements while viewing natural scenes without any particular task. To better understand the visual features that drive attention, the aim of this paper was to quantify the influence of color on eye movements when viewing dynamic natural scenes. The influence of color was investigated by comparing the eye positions of several observers eye-tracked while viewing video stimuli in two conditions: color and grayscale. The comparison was made using the dispersion between the eye positions of observers, the number of attractive regions measured with a clustering method applied to the eye positions, and by comparing eye positions to the predictions of a saliency model. The mean amplitude of saccades and the mean duration of fixations were compared as well. Globally, a slight influence of color on eye movements was measured; only the number of attractive regions for color stimuli was slightly higher than for grayscale stimuli. However, a luminance-based saliency model predicts the eye positions for color stimuli as efficiently as for grayscale stimuli.

Keywords: color information, eye movements, visual saliency, video

Introduction

Visual attention has been conceptualized in theories such as the Filter Model (Broadbent, 1958) and the Feature Integration Theory (FIT) (Treisman & Gelade, 1980). The latter is one of the most cited theories of attention, and divides the processes of attention into two stages: a pre-attentive and a focused one. According to the FIT, elementary visual features such as intensity, color and orientation are processed in parallel at the pre-attentive stage, and subsequently combined to drive the focus of attention. Based on this theory, Wolfe and colleagues introduced the Guided Search Model (GSM) and studied the elementary visual features that are involved in guiding attention using visual search tasks (Wolfe, Cave, & Franzel, 1989). These studies provided a list of the most important visual features that drive visual attention which is in accordance with the selectivity of the cortical cells

of the visual system to the features (Hubel, Wiesel, & Stryker, 1977). Both the FIT and the GSM were developed and justified through behavioural experiments using simple and artificial stimuli, without eye-tracking experiments.

Several studies have been conducted to determine the contribution of different features to the deployment of attention. Wolfe and Horowitz (Wolfe & Horowitz, 2004) classified the visual attributes when performing a visual search from undoubtedly guiding attributes—color, motion and orientation—to otherwise non-guiding attributes, such as intersection and light sources. According to that study, color is one of the most guiding attributes.

Several recent studies also investigated the role of color information in visual perception using equiluminant stimuli (Krauskopf, 1999; Hawken, Gegenfurtner, & Sharpe, 1999; Rhea & Eskew, 2009). These researches indicate that, contrary to the conclusion of early studies (Livingstone & Hubel, 1987), the color vision system is as efficient as the luminance vision system in perceiving and processing the visual information. But, the results of these studies on equiluminant stimulus could not be extended to the natural visual scenes. Additionally, the red-green color vision system evolved after the luminance vision system was already operating (Nathans, 1999; Dominy & Lucas,

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

2001). The question is why the trichromatic color vision evolved? What does color information add to the luminance information? The most common answer is to distinguish edible fruit from green one (Sumner & Mollon, 2000) or young leaves (Sumner & Mollon, 2000; Dominy & Lucas, 2001). However, psychophysical investigations show that the role of color vision might be more general. Studies on natural images show that color attributes significantly improves the recognition memory of natural scenes (Gegenfurtner & Rieger, 2000; Wichmann, Sharpe, & Gegenfurtner, 2002) and the identification of the gist of scenes (Castelhano & Henderson, 2008). Contrary to the large number of studies dealing with the importance of color for visual perception, few studies directly assess whether color influence or not visual attention through eye movements.

Visual attention and eye movements are correlated. In fact, visual attention precedes an eye movement to its goal, (Rizzolatti, Riggio, Dascola, & Umiltá, 1987; Hoffman & Subramaniam, 1995). Therefore, visual attention can be quantified via eye movement analysis when viewing complex stimuli—static natural scenes (Santella & DeCarlo, 2004; Tatler & Vincent, 2008; Bindemann, 2010; Ho-Phuoc, Guyader, & Guérin-Dugué, 2012), as well as dynamic scenes (Carmi & Itti, 2006; Dorri, Martinetz, Gegenfurtner, & Barth, 2010; Mital, Smith, Hill, & Henderson, 2010; Coutrot, Guyader, Ionescu, & Caplier, 2012).

Several computational models of attention have been developed based on the FIT and GSM (Itti, Koch, & Niebur, 1998; Itti, 2005; Frintrop, 2005; Le Meur, Le Callet, & Barba, 2007; Marat et al., 2009). These models predict regions that might be gazed while exploring natural scenes, generating saliency maps. Features such as intensity, color and spatial frequency are considered to determine the visual saliency of regions in static images, and motion is also considered in the case of dynamic scenes.

All the computational models cited above use color as a feature that drives the attention, except the model proposed by Marat and colleagues (Marat et al., 2009). The latter considers only luminance features for computing the saliency maps. Unfortunately, this model was only tested on grayscale videos. Very recently, we have shown that the incorporation of color features into this model significantly improves its performance in predicting eye positions (Hamel, Guyader, Pellerin, & Houzet, 2015). But, the video stimuli used in this previous study to evaluate the model had the specificity to include only person-present scenes.

As in computational models, in eye-tracking experiments, the influence of color on eye movements when viewing natural scenes is still being debated. Some eye-tracking studies suggest that color has very little effect (Baddeley & Tatler, 2006) or no effect on eye position, but Ho-Phuoc and colleagues report an effect on fixation duration with shorter fixations for color

images (Ho-Phuoc et al., 2012). Another study shows that the effect depends on the category of images (Frey, Honey, & Knig, 2008). Frey and colleagues investigate the saliency of different color features (saturation, red-green and yellow-blue contrasts) within seven semantic categories of images: face, flower and animal, forest, fractal, landscape, man-made, and rainforest. They report that the contribution of color features to attention depends on the category of the images. Color information increases the congruency of fixation position between participants in *rainforest*, while in *fractal* color decreases the congruency.

All these studies only address the case of static scene whereas natural scenes are mostly dynamic. In fact, motion is found to be one of the most crucial features in guiding eye movements (Itti & Baldi, 2009; Mital et al., 2010; Marat et al., 2009). Therefore, the present study aims at evaluating the contribution of color to guiding eye movements for dynamic scenes.

In this study, we compared the eye movements of different participants when viewing color videos and the same videos in grayscale, to determine whether color information influences eye movements. Quantifying the influence of color might be of interest for computational models of visual attention. Because differences were found in static images as a function of their semantic category (Frey et al., 2008), we chose videos with various contents and videos that can be classified into different categories, where color might be more or less important. We examined the effect of color, both globally and as a function of the category, on different parameters extracted from recorded eye movements: the eye positions, the duration of the fixations, and the amplitude of the saccades. The comparison was made both on average over the whole video and frame-by-frame taking into account the course over time of the video. Such a methodology was already used in a previous study analysing the influence of sound on eye movements (Coutrot et al., 2012). Finally, we measured the influence of color by comparing the eye positions recorded for color and for grayscale videos to a luminance-based saliency model (Marat et al., 2009).

Method

Participants

Thirty-seven volunteers, (17 women and 20 men, aged from 18 to 47 years, mean = 29 ± 5.5) took part in the experiment. All reported normal or corrected to normal visual acuity, while their normal color vision was tested using Ishihira color plates, presented on the experimental display. All participants gave their consent to take part in the experiments.

Stimuli

Our dataset consisted of 20 video clips, each for about 20 seconds. These clips were created by con-

catenating 134 short video snippets of from one to three seconds, called *video snippets*. We concatenated the snippets to increase the heterogeneity of the visual stimuli and to reduce possible top-down processes (Carmi & Itti, 2006; Marat et al., 2009). The snippets were extracted from various color video sources, including professional videos, such as films, TV series, and documentaries, and also amateur videos of urban roads. The stimuli had a spatial resolution of 640×480 pixels (25×19 degrees of visual angle) and a temporal resolution of 25 frames per second.

The chosen snippets were classified according to their contents into the following categories: *daylight outdoor* scenes (42 snippets), *night light outdoor* scenes (26 snippets), *indoor* scenes (37 snippets) and *urban road* scenes (29 snippets). The main difference between *urban road* and *daylight outdoor* categories was the presence of traffic signs in the former. Because traffic signs are considered particularly salient in a scene (Itti, 2005), the videos including them were considered as a separate category. Figure 1 shows some frames from each category in color and grayscale.



Figure 1. Example frames in color (first and third rows) and grayscale (second and fourth rows). The columns from left to right correspond to the categories *daylight outdoor*, *night light outdoor*, *indoor*, and *urban road*.

Initially, the videos were in different compressed formats. We converted all videos to uncompressed AVI format.

The eye-tracking experiment was setup to collect the eye movement data recorded for two stimulus conditions: color and grayscale. In order to only measure the influence of color on eye movements, we need to ensure that the luminance information is unchanged between the two stimulus conditions. But, color to grayscale conversion is a lossy operation that modifies the luminosity features of the video stimuli. Color to grayscale conversion is required in many applications such as in rendering color videos to a monochrome device, printing color documents in grayscale. It might

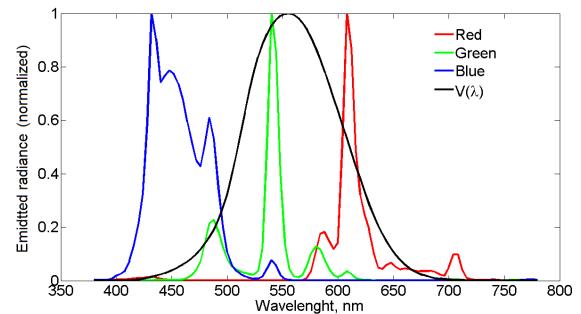


Figure 2. Spectral power distribution for light emitted by the red, green, and blue phosphors of the experimental display and the CIE 1931 luminosity function of the standard observer, $V(\lambda)$.

also be a pre-processing step in the context of vision algorithms, for example in stereo matching algorithms. According to the applications several grayscale conversion methods have been developed trying to preserve the perceptual properties of the original color image (Gooch, Olsen, Tumblin, & Gooch, 2005; Kim, Jang, Demouth, & Lee, 2009; Benedetti, Corsini, Cignoni, Callieri, & Scopigno, 2010). The NTSC conversion is maybe one of the most common methods which is extracted from the ITU recommendation 601. This method is based on a weighted sum of the R, G and B channels that considers the luminosity function of standard observer, $V(\lambda)$, and also the spectral distribution of the primaries of the display. But the weights of R, G and B channels do not correspond to all display types. Here, we used a grayscale conversion method that still corresponds to a weighted sum of the R, G and B channels, but considers the characteristics of display, Equation 1.

$$L = 0.5010 \times R + 0.4911 \times G + 0.0079 \times B \quad (1)$$

The weights of the R, G and B channels were calculated according to the experimental display characteristics to fit $V(\lambda)$, the CIE 1931 luminosity function of the standard observer. The display characteristics were obtained by measuring the light emitted from a computer controlled display using a Photo Research PR650 spectrometer. Figure 2 presents the spectral power distributions of the R, G and B channels. This conversion method is adopted to our experimental display and supports the luminance matching between the grayscale and the color versions of the stimuli.

Apparatus

An LCD color monitor of 21 inches at a refresh rate of 85 Hz was used to display the video clips. The participants were at a distance of 57 cm from the display, resulting in a visual stimulus over 25×19

degrees of visual angle. The eye movements were recorded with an SR research Eyelink 1000 eye-tracker. The eye-tracker was used in a pupil-tracking mode at a frequency of 1000 Hz. The stimulus presentation, synchronization, and recording were carried out by software developed in our laboratory (Ionescu, Guyader, & Guérin-Dugué, 2009). Only the dominant eye of each participant was tracked.

Experimental design

Each experiment session was divided into two parts. During the first part, the participants watched one-half of the video clips in one stimulus condition (color/grayscale), while during the second part, the participants watched the other half of the videos in the other condition (grayscale/color). Thereby, each color video snippet was observed by 18 subjects, and each grayscale video snippet was observed by 19 subjects. Each part started with a 9-point eye-tracker calibration. Moreover, each video clip started with a drift correction. A new calibration was run if the drift error was above 0.5 degrees. Each video was followed by a gray background displayed for 2 s. Both parts took place on the same day in a darkened room in the presence of the experimenter. The participants were asked to carefully watch the video clips while keeping their head immobile on a chin rest.

Data

During the experiment, the eye movements of the participants were recorded. The eyelink software reported, in a data file at each millisecond, the raw eye positions and some detected events, such as saccades, fixations, and blinks. We extracted the eye positions of the participants on the video frames, the duration of the fixations, and the amplitude of the saccades for each participant.

Eye Positions. For each participant, 40 raw eye positions per frame were recorded. These 40 positions were summarized into a *median position* with median *x* and median *y* coordinates, referred to as the eye position of one participant per frame. To simplify the notation, the eye positions recorded under the color stimulus condition are called *color positions* (*C*), whereas eye positions under the grayscale stimulus condition are called *grayscale positions* (*GS*).

Duration of Fixations and Amplitude of Saccade. The EyeLink 1000 tracker parser detects saccades according to three thresholds: motion (degrees), velocity (degrees/sec), and acceleration (degrees/s²). Here, the acceleration, velocity, and motion thresholds were set to 30 degrees/s, 8000 degrees/s² and 0.15 degrees, respectively. We analysed both the amplitude of the saccades and the duration of the fixations.

Eye position analysis metrics

Dispersion. To evaluate the variability of the eye positions between the participants, we used a metric called the *dispersion* (Marat et al., 2009; Salvucci & Goldberg, 2000). This metric was computed using the *leave one out* method (Torralba, Oliva, Castelhano, & Henderson, 2006). First, the Euclidean distances between the eye position of one participant and the eye positions of the other participants were calculated. Then the final dispersion for each frame was obtained by averaging the dispersion over all participants,

$$D = \sqrt{\frac{1}{N^2} \sum_{i,j < i} d_{i,j}^2} \quad (2)$$

where *N* is the number of eye positions for a frame and $d_{i,j}$ is the Euclidean distance between the eye positions of participants *i* and *j*.

The dispersion was calculated for each frame separately, for *C* positions of each frame (D_C) and *GS* positions (D_{GS}). It measures the variability between the eye positions of the participants for each stimulus condition. Lower values of the dispersion are observed when the eye positions are located in similar positions: this is interpreted as a high level of inter-participant consistency.

Clustering. The salient objects of a visual scene correspond to the regions of interest of a scene fixated by a group of participants at the same time. These regions can be estimated for each frame by clustering the recorded eye positions. Here, we clustered the eye positions to compare the number of regions of interest between the color and grayscale conditions.

Clustering methods use distance metrics between the eye positions to find the regions of interest. *K*-means is one of the clustering methods previously used to cluster eye positions (Follet, Le Meur, & Bacino, 2011; Privitera & Stark, 2000; Latimer, 1988). This method has one main drawback: the number of clusters must be determined *a priori*. Another clustering method, which leads to consistent results, is the mean-shift method. Santella and DeCarlo (Santella & DeCarlo, 2004) employed this method on eye fixations to quantify visual areas of interest. The mean-shift algorithm is a non-parametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. In this study, we employed this method to cluster the eye positions per frame. In this clustering method, a distance parameter is required. Since all video clips have the same size, we set empirically this distance to 100 pixels, equal to approximately four degrees of visual angle.

Saliency map analysis

Visual saliency models have been developed to predict the regions that have the highest probability of attracting the gaze of observers. The fixated regions are differed from non-fixated regions according to their low level features. Here, we compared the C and GS eye positions to a luminance-based computational saliency model (Marat et al., 2009).

The computational saliency model of Marat and colleagues consists of two visual pathways: *static* and *dynamic*, dedicated to different types of visual features. These pathways are only based on the luminance information, and emphasize the regions that differ from their surroundings in terms of the spatial frequencies and orientations for the static pathway, M_s , and in terms of the motion amplitude for the dynamic pathway, M_d , Figure 3.

A classical metric for comparing eye positions to a computational saliency map is the Normalized Scanpath Saliency (NSS) (Itti, 2005). We used this metric to compare the C and GS eye positions with the saliency map of the corresponding scene. To compute this metric, first the saliency maps are normalized to zero mean and unit standard deviation. The NSS value of a given frame corresponds to the average of the values of the normalized saliency map at the eye positions.

A high positive NSS value indicates that the eye positions are located on the salient regions of the computational saliency map. An NSS value close to zero represents no relation between the eye positions and the computational saliency map, while a highly negative value of NSS means that the eye positions are not located on the salient regions of the computational saliency map.

Results

The aim of this study was to determine how color influences eye movements during free viewing of videos. The main question was whether color influences the location of the gaze. The design of our experiment allowed us to compare the eye positions recorded while viewing color and grayscale stimuli. We studied the influence of color on the variability between the eye positions of the different participants using the dispersion metric. We also compared the number of regions of interest under color and grayscale conditions using the mean-shift clustering method. These two metrics, dispersion and clustering, were computed for each frame. Moreover, we compared the duration of the fixations and the amplitudes of the saccades under both conditions. Finally, we compared the eye positions under the two stimulus conditions to the computational saliency maps.

We analysed the effect of the stimulus category (day-light outdoors, night light outdoor, indoor, or urban roads) and the effect of the stimulus condition (color

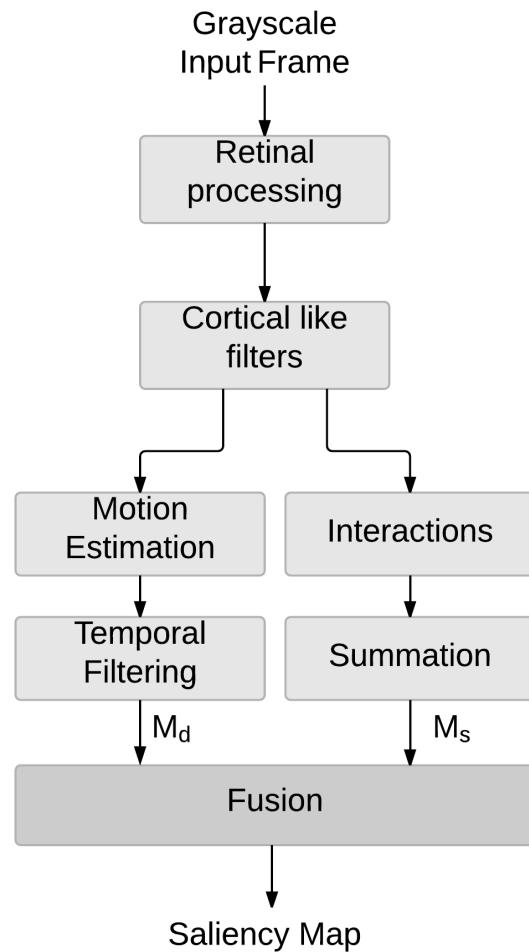


Figure 3. The spatio-temporal saliency model: M_d is the luminance-based dynamic map, M_s is the luminance-based static map (Marat et al., 2009).

or grayscale) on the different metrics obtained from the eye-tracking experiment: Dispersion, number of clusters, duration of fixations, amplitude of saccades, and NSS. All the statistical analyses were run per item (video snippets correspond to observations).

We also studied the temporal evolution of these metrics frame-by-frame. We limited the temporal analysis to the first 65 frames of each snippet, because most of the snippets have at least 65 frames and the influence of a top-down attention on the participants would be minimal this way. We defined three periods of observation: early (frames 1 to 15, 600 ms), middle (frames 16 to 40, one second) and late (frames 41 to 65, one second). The terminology is similar to that used by Follet and colleagues (Follet et al., 2011) for static images. These metrics were computed frame-by-

frame and were averaged over all frames for each video snippet.

Dispersion of eye positions

First, the dispersion was analysed on average over the whole snippet. Figure 4 shows the mean dispersion under color and grayscale stimuli according to the stimulus category. Repeated measures ANOVA were run with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (color, grayscale) as a within-item factor.

We observed a principal effect for the *Stimulus Category* ($F(3,130) = 4.09, p < 0.01$). But, no effect of the *Stimulus Condition* ($F(1,130) = 2.06, p = 0.15$), or interaction of the *Stimulus Condition* \times *Stimulus Category* ($F(1,130) = 1.28, p = 0.29$) was observed.

We ran Bonferroni multiple comparison tests to compare the mean dispersions obtained for the different categories. The mean dispersion for *night light outdoor* category was lower than those for the categories *daylight outdoor* and *indoor* ($p < 0.01$). This was expected, because in this category a limited region has been illuminated that makes observers looking only at these regions.

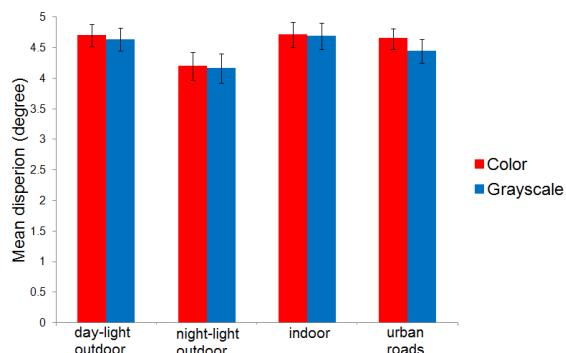


Figure 4. Mean dispersion according to stimulus category for color stimuli (red columns) and for grayscale stimuli (blue columns). Error bars represent the standard error (standard deviation divided by the root square of the number of items).

We also studied the temporal evolution of the dispersion. Figure 5 shows the evolution of the mean dispersion for the color and grayscale stimuli as a function of the viewing time (frame rank), through three periods of observation: early (frame 1 to 15), middle (frame 16 to 40) and late (frame 41 to 65). The two curves followed the same pattern for both stimulus conditions. In the early period of observation, the mean dispersion reached its minimum value (color, 3.2 ; grayscale, 3.1) and increased during the middle and the late periods. Because we did not observe any principal effect of the *Stimulus Condition* for the global analysis we did not further analyse the effect of the *Period of Observation*.

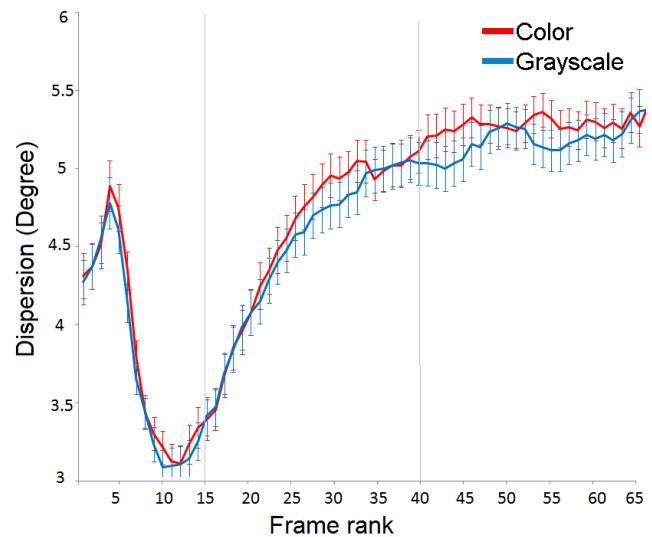


Figure 5. Mean dispersion according to the stimulus condition (color and grayscale) in degrees of visual angle across time (frame rank). Error bars represent the standard error (standard deviation divided by the root square of the number of items).

Number of clusters in eye positions.

Clustering the eye positions emphasizes the most attractive regions of scene. Figure 6 shows the mean number of clusters for color and grayscale stimuli according to stimulus category. As for dispersion, a repeated measures ANOVA was run with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (color, grayscale) as a within-item factor. A principal effect was observed for the *Stimulus Category*, $F(3,130) = 4.4; p < 0.005$, as well as for the *Stimulus Condition*, $F(1,130) = 4.9; p < 0.03$. However, no effect of the interaction of *Stimulus Condition* \times *Stimulus Category* was observed, ($F(3,130) = 0.374; ns$).

Bonferroni multiple comparison tests showed that the mean number of clusters for the *night light outdoor* category was lower than that for the *daylight outdoor* and *indoor* categories ($p < 0.01$). This result reinforced the previous result on the smaller dispersion for this category.

Contrary to the dispersion metric, clustering metric showed a significant effect of the color. The mean number of clusters for color stimuli was higher than for grayscale (1.62 versus 1.58). Even the effect of color was small regarding the mean number of clusters, it might reveal that color increases the number of fixated regions and hence the number of salient regions. Figure 7 shows an example frame with the regions of interest for color (red ellipses) and those for grayscale (green ellipses).

Finally, we analysed the temporal evolution of the

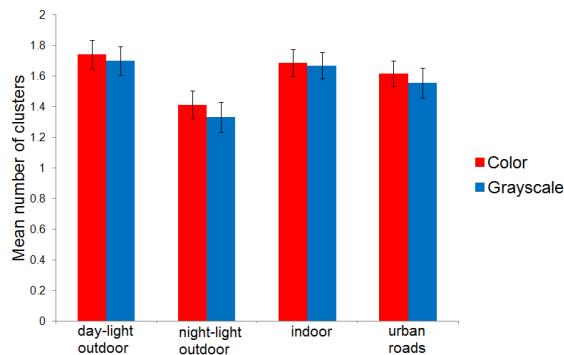


Figure 6. Mean number of clusters according to stimulus category. Error bars represent the standard error (standard deviation divided by the root square of the number of items).



Figure 7. An example scene depicting the different clusters. The ellipses depict the clusters of eye positions obtained from the mean-shift clustering method. Red ellipses represent the clusters extracted from color eye positions and green ellipses represent the clusters extracted from grayscale eye positions.

mean number of clusters, Figure 8. We ran repeated measures ANOVA with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (color, grayscale) and *Period of Observation* (early, middle and late) as within-item factors. We observed a principal effect of the *Stimulus Condition* ($F(1,112) = 9.7; p < 0.001$), a principal effect of the *Period of Observation* ($F(2,224) = 2.46; p < 0.001$), and a principal effect of the *Stimulus Category* ($F(3,112) = 2.9; p < 0.05$). A significant effect of the interaction of the *Stimulus Condition* \times *Period of Observation* was also observed, ($F(2,224) = 14.5; p < 0.0001$). Finally, no effect of the triple in-

teraction was observed. As shown in Figure 8, in the early period of observation there is no significant difference between the mean number of clusters for color and grayscale stimuli. But, in the middle period of observation, the mean number of clusters for color stimuli is higher than that for grayscale (1.67 vs 1.61), and this effect persists in the late period of observation (1.86 vs 1.82).

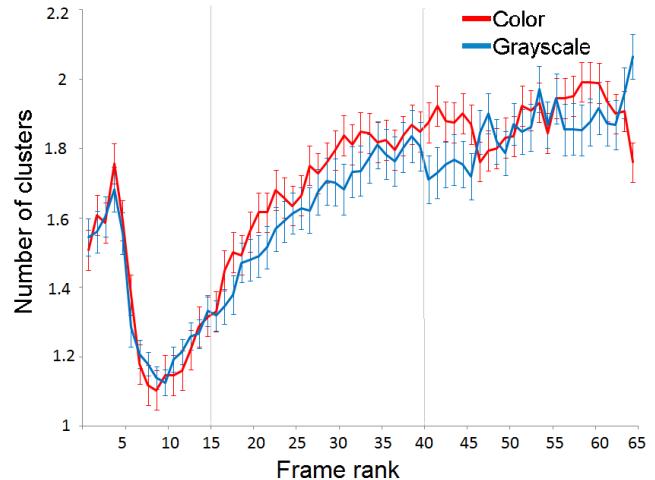


Figure 8. Mean number of clusters according to the stimulus condition over time (frame rank). Error bars represent the standard error (standard deviation divided by the root square of the number of items).

Duration of fixations and amplitude of saccades

To assess the influence of color information on eye movements, we also studied the duration of the fixations and the amplitude of the saccades. Two separate repeated measures ANOVA were run with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (color, grayscale) as a within-item factor.

For the mean duration of the fixations, a principal effect of *Stimulus Category* ($F(3,130) = 11.71, p < 0.001$) was observed. But, we observed no effect of *Stimulus Condition* (color, 318 ms versus grayscale, 324 ms, $F(1,130) = 0.36, p = 0.55$), or of the interaction of *Stimulus Condition* \times *Stimulus Category* ($F(1,130) = 0.52, p = 0.68$). Bonferroni multiple comparisons were run to determine which categories were different from the other categories. The mean duration of the fixations for the *night light outdoor* category was higher than for the other three categories (*night light outdoor*: 373 ms versus *daylight outdoor*: 307, *indoor*: 290 and *urban roads*: 314 ms, $p < 0.01$).

We also observed a principal effect of *Stimulus Category* on the amplitudes of the saccades, (*night light outdoor*: 3.89, *daylight outdoor*: 4.52, *indoor*: 4.41 and *urban roads*: 4.52 degrees, $F(3,130) = 11.71, p < 0.001$). But, no effect of *Stimulus Condition* (color, 4.35 degrees

versus grayscale, 4.41 degrees $F(1,130) = 0.36, p = 0.55$), or of the interaction of *Stimulus Condition* \times *Stimulus Category* ($F(1,130) = 0.52, p = 0.68$) was observed.

Bonferroni multiple comparisons determined that the mean amplitude of the saccades for *night light outdoor* category is higher than for *daylight outdoor* (night light outdoor: 3.9 degrees, daylight outdoor: 4.52 degrees, $p < 0.05$).

In summary, the results show that the stimulus categories that were used in this experiment do not influence eye movements, except for the *night light outdoor* category. Independent from the stimulus condition, for *night light outdoor*, we observed that the mean dispersion and mean number of clusters are lower than other categories, the mean duration of fixations is higher and the mean amplitude of saccades is lower. These results might be because of the particular composition of the *night light outdoor* category, where only a limited part of the scene is illuminated. The analysis of the eye positions recorded under the free viewing of color and grayscale videos shows that the color information influences neither the between-participant congruency (the inter-participant dispersion) nor the amplitudes of the saccades nor the duration of the fixations.

The clustering of the eye positions indicates that color information increases the number of fixated regions, which suggests that color information makes some new regions salient.

Saliency model

We studied the performance of a luminance-based saliency model, the one proposed by Marat and colleagues (Marat et al., 2009), to predict the two data sets of eye positions. If the prediction efficiency of the model for both color and grayscale eye positions is similar, we might conclude that only the luminance information is necessary to predict the gazed regions even for color stimuli.

The two datasets of eye positions C and GS were compared to the saliency maps of the model of Marat and colleagues (Marat et al., 2009), Figure 3. The NSS score was used to compare the C and GS positions to the saliency maps. The NSS score for color and grayscale eye positions is similar (0.89 versus 0.91). We did not observe any effect of *Stimulus Category* ($F(1,132) = 1.60, p = 0.19$) neither of *Stimulus Condition* ($F(1,132) = 2.47, p = 0.12$). No effect of the interaction of the two factors was observed, either ($F(1,132) = 1.72, p = 0.19$).

Discussion

In this study, we measured the influence of color information on the eye movements recorded during the free exploration of videos. We compared the eye positions for color and grayscale stimuli. We used

a display-dependent grayscale conversion method to ensure the luminance matching between color and grayscale stimuli. The grayscale version of stimuli were obtained from the weighted sum of color channels to fit $V(\lambda)$. However, this conversion method is still lossy and the $V(\lambda)$ corresponds to the average standard observer while the response of photo-cells varies from one observer to another and the random cone mosaic of human eye might affect equiluminance thresholds (Alleysson & Meary, 2012).

Color and grayscale eye positions were compared using various metrics: the dispersion and the mean number of clusters to directly compare the eye positions, the mean amplitude of the saccades, the mean duration of the fixations, and finally, the similarity of the eye positions to the predictions of a saliency model. All the comparisons were also done taking into account the semantic category of the dynamic scene. We studied different categories : *daylight outdoor*, *night light outdoor*, *indoor*, and *urban roads*. Evidences from research of Frey and colleagues (Frey et al., 2008) show that the influence of color on eye positions depends on the semantic category of the image. The latter study introduced two extreme categories of static images: *fractal* and *rainforest*. In *fractal*, color information renders the participants' fixation patterns more dissimilar, whereas in the *rainforest* category, color increases the participants' consistency significantly. Based on the conclusions of that study, we had anticipated that the influence of color on eye positions would be related to the category of the video snippet. Here, we instead found that the influence of color remains insignificant across different categories of videos. Concerning the influence of category, independent from the stimulus condition, we found that for videos belonging to the *night light outdoor* category eye movements are different from the ones for the other categories.

Concerning the effect of stimulus condition, we found that color does not influence the dispersion metric, i.e., the variability of the eye positions among participants. Yet, the number of clusters of the eye positions showed that there are slightly more clusters for color eye positions than for grayscale eye positions. These results might suggest that color information increases to certain extend the number of salient regions in the dynamic scenes. Moreover this effect was not constant across the periods of viewing time being larger in the middle period (frame 16 to 40).

The temporal analysis of eye positions showed a typical shape for the evolution of the mean dispersion and the mean number of clusters according to the frame rank. Note that this evolution is independent of the stimulus condition. In the early period of observation, eye positions are influenced by the central bias (Tatler, 2007; Bindemann, 2010; Marat, Rahman, Pellerin, Guyader, & Houzet, 2013). This could be observed on the two curves of figures 5 and 8. Due to this bias, a high consistency of the eye positions of

participants is observed about 400 ms (the 10th frame) after the onset of a stimulus, which is in accordance with the low dispersion, as well as the small number of clusters for color and grayscale eye positions. Then both metrics increase to reach a plateau.

In addition, for dynamic scenes, we found that color information does not influence the duration of fixations neither the amplitude of saccades; this result differs from a previous study on static images (Ho-Phuoc et al., 2012). This difference between static and dynamic scenes, concerning the influence of color on eye movements, could be due to the temporal changes and dynamic nature of the video stimuli. Moreover, the viewing time in the present experiment is shorter than those for the mentioned experiments with static images (Ho-Phuoc 5 sec, Frey 6 sec, present study 2 to 3 sec depending on the duration of the stimulus).

Finally, we compared the two data sets of eye positions, recorded for color and grayscale videos, to a saliency model. The luminance-based saliency model, initially developed by Marat and colleagues (Marat et al., 2009) has similar prediction efficiency for color and grayscale stimuli. Therefore, a saliency model simply based on luminance information is efficient to predict eye positions recorded for color video stimuli. Note that this main result differs from our previous study for which we found that incorporation of color information to the luminance-based saliency model proposed by Marat, significantly improves the performance of the model (Hamel et al., 2015). However these different results might be explained by the fact that in the previous study we used very specific stimuli depicting only person-present scenes. In this new experiment we used video stimuli with more various content. Future experiments might generalize these results to a larger database.

To conclude, the results of present experiment do not reveal a significant influence of color information on the eye movements when exploring natural video stimuli, even a slight effect of color on the mean number of clusters is found (with a significant effect in the middle period of viewing time). These observations might suggest that color features might have a small contribution in performance of saliency models, at least for models that predict the gazed regions in videos with various contents.

References

- Alleysson, D., & Meary, D. (2012). Neurogeometry of color vision. *Journal Of Physiology Paris*, 106, 284-296.
- Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, 46(18), 2824-2833.
- Benedetti, L., Corsini, M., Cignoni, P., Callieri, M., & Scopigno, R. (2010). Color to gray conversions in the context of stereo matching algorithms. *Machine Vision and Applications*, 57(2), 254-348.
- Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, 50(23), 2577-2587.
- Broadbent, D. E. (1958). *Perception and communication*. (J. E. Birren & K. W. Schaie, Eds.). Pergamon Press.
- Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46, 4333-4345.
- Castelhano, M. S., & Henderson, J. M. (2008). The Influence of Color on the Perception of Scene Gist. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 660-675.
- Coutrot, A., Guyader, N., Ionescu, G., & Caplier, A. (2012). Influence of soundtrack on eye movements during video exploration. *Journal of Eye Movement Research*, 5(4), 1-10.
- Dominy, N. J., & Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature*, 410, 363-366.
- Dorr, M., Martinetz, T., Gegenfurtner, K., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10), 1-17.
- Follet, B., Le Meur, O., & Baccino, T. (2011). New insights on ambient and focal visual fixations using an automatic classification algorithm. *iPerception*, 2(6), 592-610.
- Frey, H. P., Honey, C., & Knig, P. (2008). What's color got to do with it? The influence of color on visual attention in different categories. *Journal of Vision*, 11(3), 1-15.
- Frintrop, S. (2005). *VOCUS: A Visual Attention System for Object Detection and Goal-directed search* (Vol. 3899 / 2006). Springer Berlin/Heidelberg.
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, 10, 805-808.
- Gooch, A., Olsen, S., Tumblin, J., & Gooch, B. (2005). Color2gray: salience-preserving color removal. *ACM Transactions on Graphics*, 24 Publisher:(3), 1-6.
- Hamel, S., Guyader, N., Pellerin, D., & Houzet, D. (2015). Contribution of color in saliency model for videos. *Signal, Image and Video Processing*. doi: 10.1007/S11760-015-0765-5
- Hawken, M. J., Gegenfurtner, K. R., & Sharpe, L. (1999). Color Vision: From Genes to Perception. In K. R. Gegenfurtner & L. Sharpe (Eds.), (p. 283-299). Cambridge Univ. Press.
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57 (6), 787-795.
- Ho-Phuoc, T., Guyader, N., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, 12(2), 1-13.
- Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1977). Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature*, 269, 328 - 330.
- Ionescu, G., Guyader, N., & Guérin-Dugué, A. (2009). *SoftEye software*. IDDN.FR.001.200017.0000.S.P.2010.003.31235.
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12, 1093 -1123.
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295-1306.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE*

- Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254–1259.
- Kim, Y., Jang, C., Demouth, J., & Lee, S. (2009). Robust colortogray via nonlinear global mapping. *ACM Transactions on Graphics*, 14.
- Krauskopf, J. (1999). Color Vision: From Genes to Perception. In K. R. Gegenfurtner & L. Sharpe (Eds.), (p. 303-316). Cambridge Univ. Press.
- Latimer, C. R. (1988). Eye-movement data: cumulative fixation time and cluster-analysis. *Behavior Research Methods, Instruments, & Computers*, 20(5), 437-470.
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, 47(19), 2483-2498.
- Livingstone, M., & Hubel, D. H. (1987). Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement, and Depth. *The Journal of Neuroscience*, 7, 3416-3468.
- Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., & Guérin-Dugué, A. (2009). Modelling spatio-temporal saliency to predict gaze direction for short videos. *International Journal of Computer Vision*, 82(3), 231-243.
- Marat, S., Rahman, A., Pellerin, D., Guyader, N., & Houzet, D. (2013). Improving Visual Saliency by Adding Face Feature Map and Center Bias. *Cognitive Computation*, 5(1), 63-75.
- Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2010). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, 3(1), 5–24.
- Nathans, J. (1999). The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron*, 24, 299312.
- Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Trans Pattern Anal Mach Intell*, 22(9), 970 - 82.
- Rhea, T., & Eskew, J. (2009). Higher order color mechanisms: A critical review. *Vision Research*, 49, 2686-2704.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25, 31-40.
- Salvucci, D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. Duchowski (Ed.), *ETRA 2000, Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71–78). ACM.
- Santella, A., & DeCarlo, D. (2004). Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest. In *ETRA 2004, Proceedings of the 2004 symposium on Eye tracking research & applications* (p. 27-34). ACM.
- Sumner, P., & Mollon, J. D. (2000). Catarrhine photopigments are optimized for detecting targets against a foliage background. *Journal of Experimental Biology*, 203, 19631986.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7, 41-17.
- Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2):5, 1-18.
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006, October). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.*, 113(4), 766–786.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). sensory and cognitive contributions of color to the recognition of natural scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 509–520.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, 15, 419-433.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), 495-501.