THE ROLE OF VISUAL ATTENTION IN THE AESTHETIC APPEAL OF CONSUMER IMAGES: A PRELIMINARY STUDY

Judith A. Redi, Isabel Povoa

Intelligent Systems, Delft University of Technology, The Netherlands
j.a.redi@tudelft.nl, misabelpovoa@gmail.com

ABSTRACT

Predicting the aesthetic appeal of images is of great interest for a number of applications, from image retrieval to visual quality optimization. In this paper, we report a preliminary study on the relationship between visual attention deployment and aesthetic appeal judgment. In particular, we seek to validate through a scientific approach those simplicity and compositional rules of thumb that have been applied by photographers and modeled by computer vision scientists in computational aesthetics algorithms. Our results provide a confirmation that both simplicity and composition matter for aesthetic appeal of images, and indicate effective ways to compute them directly from the saliency distribution of an image.

***Index Terms*—** Aesthetic appeal, visual attention, visual quality, image saliency

1. INTRODUCTION

The possibility to predict the aesthetic appeal of images has recently attracted a lot of interest from the multimedia community, being it crucial for a number of applications, from multimedia information retrieval to computer graphics [1]. Recent research has also shown that aesthetic appeal of images plays a role in the tolerance that users have to visual distortions [2]. Augmenting objective quality metrics [3] with a prediction of aesthetic appeal could therefore significantly improve their ability to assess the overall pleasantness of images towards a finer optimization of multimedia delivery systems.

Computational aesthetics models [1] have attempted to mimic processes underlying the human appreciation for image aesthetics. Factors such as color rendering [4], semantic content [5], familiarity [6], image simplicity [7] and compliance to compositional rules [8] have been modeled through computer vision techniques towards a reliable estimation of aesthetic appeal. Existing models have achieved good performance, but the room for improvement is still large. This might be due to the fact that most of these models are often inspired by photographers' rules-of-thumb [6], which have not been validated in a scientific way. Both image simplicity (i.e., clarity of the subject [7]) and compositional rules (e.g., the well-known rule of thirds [8]), for example, are tools used by photographers to guide the observer's visual attention towards the image subject and ease perceptual fluency. Very few studies have attempted at checking their validity in a systematic way, e.g. by verifying with empirical measurements the existence of a relationship between the deployment of visual attention, image simplicity/compliance to compositional rules and aesthetic appeal. In fact, several studies have looked into the relationship between visual attention and art [9, 10]; however, they mostly analyzed visual scan paths, in relation to either viewing task [9] or painting genre [10]. To the best of the authors' knowledge, no studies so far have inquired the role of visual attention in aesthetic appeal in relation to composition or image simplicity for regular consumer images.

On the other hand, studying visual attention in relation to image preferences has been shown to have a high added value in the contingent field of objective image quality assessment [11]. The deployment of visual attention was shown to play a major role in quality appreciation: artifacts visible in the region of interest of an image are more likely to be noticed and therefore more annoying for observers [12]. As a result, modulating the distortion visibility measurements with saliency information was shown to be beneficial for objective metrics' accuracy. Also, it has been shown that in some cases, it might be sufficient to compute distortion visibility only in the region of interest of the image, implying significant savings in terms of computational complexity of the metrics [11]. Similar principles could be applied to computational aesthetics metrics; however, until now little work has been done in this direction, besides several, remarkable attempts at estimating compliance to compositional rules through the use of visual attention models [13, 14].

In this work, we present the (preliminary) findings of a large study involving 200 consumer images and 14 participants, whose eye movements were tracked while judging the aesthetic appeal of the images. We analyze to what extent the way visual attention is deployed during the image evaluation is related to simplicity, composition, and eventual aesthetic appeal. To achieve this, we define several indicators of attention deployment based on fixation and saliency [15] information. We confirm that simplicity (in

Fig. 1. Samples from the image database used in the experiment, along with their categories.

terms of low image clutter [7]) is positively correlated to aesthetic appeal, and that compositional rules such as the rule of thirds can be validated through saliency analysis. Furthermore, the indicators we define for simplicity and composition analysis can be easily implemented in computational aesthetics metrics starting from the output of visual attention models (e.g., [16]).

2. AN EYE TRACKING STUDY FOR UNDERSTANDING AESTHETIC APPEAL

We designed a within-subjects experiment, in which fourteen participants were asked to judge the aesthetic appeal of images while their eye movements were being tracked. The number of participants was chosen in line with what advised in [17, 18]. Previous studies based on eye-tracking also have shown that a number around 15 is sufficient to guarantee stable results [2,12].

2.1. Image material

A set of 200 images was included in the experiment. Of these, 56 corresponded to those already included in study [2], 26 were chosen from images freely available online, and 118 were taken from the private collection of an amateur photographer.
Images were selected to cover a wide range of subject categories, keeping the sample as representative as possible of a general image population. The dataset was labeled based on 16 categories from the website 500px.com, for both expert and amateur photography (see Fig. 1). The following criteria were considered when selecting the categories:

- Compliance to categories used in computer vision literature (e.g., the LHI dataset [19]), as in the case of *Landscapes*, *People* and *Sport*.
- Frequent occurrence in social networks, as in the case of *Food* and *Fashion*.
- Need to encompass different levels of familiarity [6], as in the case of *Abstract* and *Celebrities*.

2.2. Apparatus

The experiment was performed in a room with constant illumination at approximately 70 lux, in an environment compliant to ITU recommendations [18]. A 23" LED backlight monitor having a resolution of 1360x768 was used to display the stimuli. Participant's face movements were constrained by a chinrest at a distance of 0.7 meters from the display. A *SensoMotoric Instruments* GmbH Eye Tracker with a sampling rate of 50/60 was used to track the participants' eye movements during the image viewing. The instrument has a pupil tracking resolution of 0.1° and a gaze position accuracy of 0.5 to 1.

2.3. Methodology

For each image in the database, participants were asked to score its aesthetic appeal in a Single Stimulus setup [18], using a 5-point discrete scale ranging between very low (1) and very high aesthetic appeal (5).
Because of the large number of images involved, fatigue and memory effects might have affected the data collection (as revealed by a pilot experiment). As a consequence, participants were asked to score images in two sessions, involving 100 images each, to be performed in different days. Each session lasted on average 40 minutes, including a short break after scoring the first 50 images.
All participants were first briefed about the general setup of the experiment and their task. A short training session (consisting in rating 3 images) was performed to allow participants to familiarize with their task. The images provided in the training were not intended to be anchoring stimuli for the scoring scale, as we did not want to prime participants with specific criteria for judging images. Participants had no time constraints in observing the images prior to scoring (both in the training and in the actual experiment). Before each image, participants' initial fixation point was forced to be in the center of the image by displaying a white cross in the middle of the screen (with a neutral background). The scoring scale was accessible only after completing the viewing of an image, in order to avoid distraction during the image observation. Images were presented in a randomized order for every participant.
At the beginning of every session (and after every break) the eye-tracker was calibrated on the participant's gaze based on a 13-points grid.

3. DATA ANALYSIS

Individual Aesthetic Appeal Scores were processed according to the procedure recommended in [18], which pointed out the presence of one outlier participant, then excluded from the analysis. Scores were then normalized per participant and transformed into individual Aesthetic Appeal Z-Scores (AAZ), eventually ranging between -3.01 and 3.47. To verify inter-observer consistency in scoring, we computed the standard deviation across the scores given to the same image by the participants, and then averaged it across all image. This resulted into a value of 0.82, corresponding to 12% of the aesthetic appeal range covered by the AAZ, in line with previous results in the field [2].

3.1. Eye tracking data analysis

We processed eye-tracking recordings in order to collect information on both eye movements and attention deployment. With respect to the latter, we processed fixation data according to [20] to obtain, per each image, visual importance information in the form of saliency maps. Saliency maps [15] represent the probability, pixel per pixel, for a location in the image to be attended by the (average) observer. As such, they outline the areas in the image which attract most attention. We believe this information can be helpful in our analysis for two main reasons. First, they may provide a powerful tool to estimate simplicity, in terms of how visually crowded (how many areas of the image attract attention, as a measure of clutter, or low simplicity) is the image. Second, it is commonly assumed that highly salient areas correspond to the most important elements in the image. Photographers intentionally compose images so that visual attention is driven to these elements; an analysis of salience could therefore reveal the compliance of an image to compositional rules-of thumb, to be later matched to an actual benefit in terms of aesthetic appeal.

The following steps were performed to create saliency maps from raw eye-tracking data:

1. All fixations lasting less than 100 ms were discarded from the recordings;
2. For each image I of size W_I x H_I, locations fixated by every observer were identified and added to a fixation map $FM^{(I)}(x,y)$, eventually gathering all fixation points from all observers;
3. $FM^{(I)}(x,y)$, was then smoothed by applying a grey scale patch with Gaussian intensity distribution whose variance (σ) was approximating the size of the fovea (~2° of visual angle). The resulting saliency map element $SM^{(I)}(k,l)$, at location (k,l) was therefore computed as:

$$SM^{(I)}(k,l) = \sum_{f=1}^{N_f} \exp\left[-\frac{(x_f-k)^2+(y_f-l)^2}{\sigma^2}\right] \quad (1)$$

with (x_j, y_j) being the pixel coordinates of the fth fixation ($f=1\ldots N_f$) in $FM^{(I)}(x,y)$, and $k\in[1, W_I]$, $l\in[1, H_I]$.

We also produced binary versions of the saliency maps SM, in order to isolate the Region(s) of Interest (ROI) of the image. To compute our Binary Maps (BM) we performed the following extra steps:

4. A saliency threshold th^S was determined, common for all maps, as one third of the maximum saliency value across all maps. A threshold for saliency was preferred over a threshold for the size of the ROI area (as used in other works, e.g., [12]), in order to isolate areas that were equally salient across all images. Of course, the value of the threshold was established in a somewhat arbitrary way and changes in the threshold may affect the results reported in the following section. We delegate to future studies further investigations on these aspects.
5. For each image I, its binary map $BM^{(I)}$ was determined as:

$$BM^{(I)}(x,y) = \begin{cases} 1 & if \quad SM^{(I)} > th^S \\ 0 & otherwise \end{cases} \quad (2)$$

4. RESULTS

4.1. Analysis of viewing strategy

As a first step, we investigated possible relationships between eye movements' characteristics and aesthetic appeal z-scores (AAZ). In particular, per each subject and image, we considered the number of fixations and saccades, their average duration, and the amplitude and velocity of saccades. These indicators are often used to describe visual strategy [21] and were found to be related to both viewing task [22] and perceived quality [23]. We report in Table 1 their mean across all observers and images, and related Standard Error. The average number and duration of fixations was found to be comparable to that obtained for other studies in the field [10] and slightly lower than that found, e.g., for technical quality scoring [21].
To check whether a relationship existed between viewing strategy parameters and judgments of aesthetic appeal, we also computed the linear correlation coefficient between these quantities and the corresponding AAZ. As shown in Table 1, none of the parameters was found to be a predictor to aesthetic appeal, as instead was found in, e.g., [22]. In

Table 1. Statistics of eye movements and correlation with aesthetic appeal judgments.

	Mean		Correlation with AAZ
	Statistic	Std. Error	
Number of fixations	20.17	0.288	-0.017
Duration of fixations	381,79	2.522	0.019
Number of saccades	16,55	0.251	-0.021
Duration of saccades	30,73	0.276	0.012
Amplitude of saccades	2,11	0.054	-0.050
Velocity of saccades	63,57	1.052	0.040

that case, the duration of fixations was negatively correlated with video quality, perhaps because the sudden appearance of visual artifacts would capture and hold attention in an unnatural way. In the case of static images, such surprise effect does not apply. This, along with difference in viewing task [22] could partially explain this discrepancy in viewing behavior.

4.2. Analysis of visual attention deployment

As mentioned in section 3.1, visual saliency can reveal important properties of the image, in particular related to visual clutter [7] and composition [8].
In the following, we describe a set of indicators that we designed to characterize both elements starting from Fixation and Saliency information and we check their relationship with aesthetic appeal.

4.2.1. Simplicity and clutter indicators

The following indicators were designed to attempt at estimating visual clutter from saliency information:

Peak Saliency: The peak value of a saliency distribution represents the location of the image that is more likely to attract the attention of an (average) observer. A high peak value indicates that in the image there is one location (i.e., an image element, possibly the subject) that is highly attractive. Lower values would instead indicate poor attractiveness, perhaps because of the presence of multiple attractive elements in the image (visual clutter). We calculate this quantity as $Peak\_S^{(I)} = max(SM^{(I)})$, and we expect it to be positively correlated to aesthetic appeal.
Saliency Spread: the spread of saliency values across the image measures whether the attention was directed towards a concentrated area (low clutter) or was instead distributed throughout the image (high clutter). We measure it by computing the standard deviation of the saliency distribution of each image: $Spread\_S^{(I)} = stdev(SM^{(I)})$, and we expect it to be negatively correlated to aesthetic appeal.
Number of fixations within the ROI: The less fixations are scattered in the background of the image, the more it is likely that there is a single object attracting the viewer's attention, which implies visual simplicity and low clutter. We compute this feature as:

$$nFix\_ROI^{(I)} = \sum_{x=1}^{W_I}\sum_{y=1}^{H_I} FM^{(I)}(x,y)\cdot BM^{(I)}(x,y) \qquad (3)$$

with $FM^{(I)}(x,y)$ being the fixation maps, and $BM^{(I)}(x,y)$ the binary map for each image I.
Dispersion of the fixations: Introduced in [23], this indicator intends to measure the spread of the fixations during the observation of an image. $Disp\_Fix^{(I)}$ is computed as the average Euclidean distance between each fixation in $FM^{(I)}$ and the centroid of fixations.
Number of distinct ROIs: When attention is divided over different elements in the image, there might be multiple peaks in the saliency distribution, and, as a result of the thresholding procedure described in section 3.1, this may originate multiple Regions of Interest in the binary maps. We define $No\_ROI^{(I)}$ the number of distinct ROIs retrievable in $BM^{(I)}$, and we expect this indicator to be negatively correlated with aesthetic appeal.

4.2.2. Indicators of compliance to composition rules

Several studies have attempted at using saliency information generated by visual attention models [13, 14] in order to predict the compliance of the image content to composition rules such as the *rule of thirds*. Such rule, often used by professional photographers, states that to ensure ease of view, the center of the main object should be located along the intersections of the lines that divide the image in thirds (see figure 2).
We replicate here a set of indicators that have been previously used for computational aesthetic models, attempting at verifying the compliance of the image to the Rule of Thirds:

- **Minimum Euclidean Distance** ($Dist\_thirds^{(I)}$) between the centroid of the (largest) ROI and the intersections of the line of thirds, normalized by the size of the image (as per [4])
- **Minimum Distance from the horizontal lines of thirds** ($Dist\_thirds\_h^{(I)}$) of the centroid of the ROI, normalized by the height of the image H_I
- **Minimum Distance from the vertical lines of thirds** ($Dist\_thirds\_v^{(I)}$) of the centroid of the ROI, normalized by the width of the image W_I

Furthermore, we compute the extent of the **area of the ROI** ($Area\_ROI^{(I)}$), normalized by the whole image area, to estimate the balance between main element and background.

4.2.3. Results

To better appreciate the impact of our indicators on aesthetic appeal, we first quantized all their values (except for indicator *No_ROI*) into three classes (low, medium and high indicator value). This was achieved by (1) detecting the 33rd and 66th percentiles of the distribution of the indicators throughout images and (2) assigning to all the images with

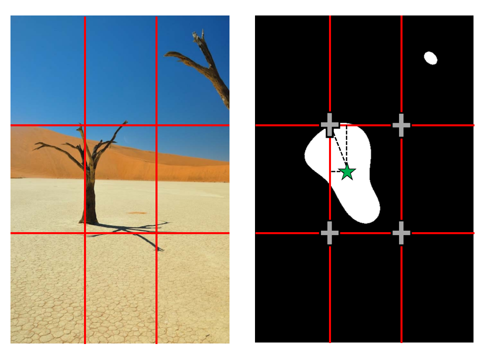

Fig. 2. Graphical explanation of the rule of thirds.

an indicator value lower than the 33rd percentile a value of 1 (low), to all images with an indicator value between the 33rd and 66th percentiles a value of 2 (medium) and a value of 3 (high) to all other images. Values of the percentiles are reported in table 2; impact of all indicators on aesthetic appeal can be visualized in fig. 3 and 4.

Figure 3 shows how our data confirm the negative effect of visual clutter on aesthetic appeal. Indicators *Peak_S* (df = 2, F = 20.88, sig = 0.000), *Spread_S* (df = 2, F = 7.38, sig = 0.001), and *No_ROI* (df = 4, F = 7.58, sig = 0.000) were found to have a significant effect on the aesthetic appeal judgments (AAZ). In particular, Figure 3.a confirms the expected relationship between *Peak_S* and AAZ: the higher the attractiveness of a single location in an image, the higher the aesthetic appeal. The relationship expected between the number of ROIs and AAZ is also confirmed (figure 3.e), with the aesthetic appeal decreasing with the increase in number of distinct ROIs (and consequent increase of clutter). An interesting effect is found for images for which no ROI was segmented (leftmost bar in Fig 3.e, *No_ROI* = 0): in this case the aesthetic appeal is also very low. This phenomenon is in line with what we expected: since we used a single threshold across all images, if no ROI was detected that was because no area in the image was sufficiently attractive to match the overall threshold. This might be due to the fact that attention was very spread across the image, which in turn could result from a high clutter in the image.

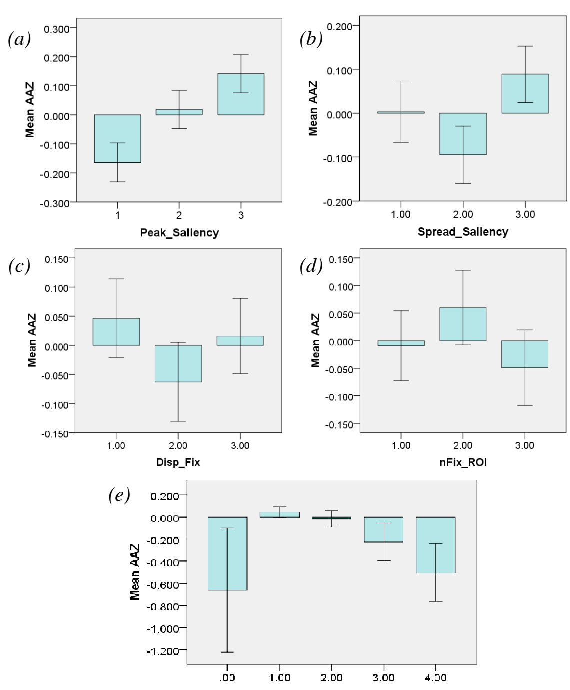

Fig. 3. Impact of clutter indicators on aesthetic appeal scores (AAZ). A level of 1 indicates low indicator values, 2 medium indicator values, and 3 high indicator values.

Figure 4 reports the relationship between the compositional features and the aesthetic appeal. We found no significant effect (df = 2, F = 2.38, sig = 0.093) of *dist_thirds* on AAZ. This is a quite interesting result, as it suggest that one of the most commonly used features to express image compliance to compositional rules might not properly reflect aesthetic appeal appreciation mechanisms. Interestingly, also features *dist_thirds_h* (df = 2, F = 1.728, sig = 0.178) and *area_ROI* (df = 2, F = 0.035, sig = 0.716) did not have a significant effect on aesthetic appeal. Conversely, the distance of the centroid of the ROI from the vertical lines of thirds *dist_thirds_v* has a significant effect on aesthetic appeal (df = 2, F = 17.32, sig = 0.000). It is also interesting to analyze the nature of this effect (figure 4.b). It seems that, for low values of *dist_thirds_v* (that is, the centroid of the ROI is close to the vertical lines of thirds) higher values of aesthetic appeal are obtained; for medium distances, aesthetic appeal significantly decreases, as expected; for high values of distance, the aesthetic appeal slightly increases again. This behavior can be explained by looking at the thresholds used to quantize the values of *dist_thirds_v* (table 2). These values are normalized by the width of the image; therefore, the maximum value that the indicator could assume is 1/3. As we can see, the maximum value found for indicator *dist_thirds_v* is ~1/6 = 0.17, which implies that in no case the centroid of the ROI is located in peripheral regions of the image. Furthermore, our analysis revealed that for the most part, ROI centroids are located in the central region of the image, i.e., that delimited by the four lines of thirds. As a result, we can assume that most of the images having a high distance of the ROI centroid from the vertical lines of thirds, are images whose ROI is located in the very center of the image. Centrality of the main subject has also been shown to be positively correlated to aesthetic appeal [7], which may partially explain our result.

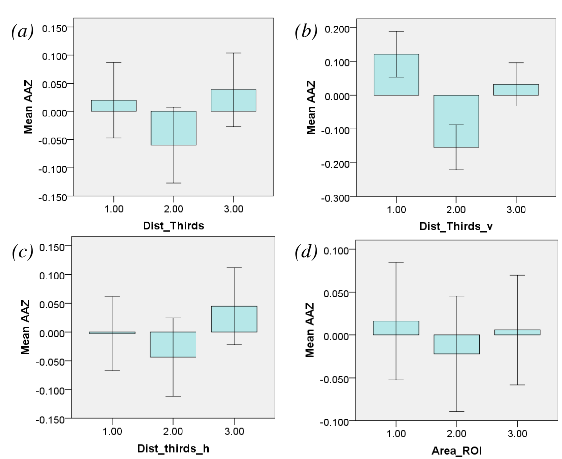

Fig. 4. Impact of image composition indicators on aesthetic appeal scores (AAZ). A level of 1 indicates low indicator values, 2 medium indicator values, and 3 high indicator values.

Table 2. Threshold used for the quantization of the indicators.

Indicator	Min	33rd Percentile	66th Percentile	Max
Peak_S	18,15	35,64	43,58	75,38
Spread_S	4,32	6,88	7,92	11,78
Area_ROI	0,00	0,05	0,08	0,28
nFix_ROI	0,17	0,52	0,64	1,00
Disp_Fix	0,27	0,41	0,50	0,77
Dist_Thirds	0,00	0,12	0,16	0,27
Dist_Thirds_v	0,00	0,10	0,14	0,17
Dist_Thirds_h	0,00	0,07	0,12	0,17

5. CONCLUSIONS

In this paper, we conducted a preliminary study on the role of visual attention in image aesthetic appeal appreciation. We tracked the eye movements of 14 subjects during the judgment of aesthetic appeal of a set of 200 consumer images and then analyzed the relationship between the attention deployment and the aesthetic appeal judgments. We designed a set of indicators extracted from human saliency (easily adaptable to saliency information gathered from computational models) that validated a negative correlation between image clutter (low simplicity) and aesthetic and a clear human preference for images having the most attractive object located either at the vertical line of thirds or the center of the image. It should also be mentioned that the influence on perceived quality of participant expertise as well as of the image content can be highly relevant, and will be investigated at a later stage.

No impact of the vertical placement of the ROI was found instead, which is useful information to simplify the computation of composition features in computational aesthetics model.

We intend to further investigate in the future on other relationships between attention deployment and aesthetics, features such as contrast of image, color and texture inside and outside the ROI. Furthermore, in the future developments of this study, we intend to validate the current findings into actual computational aesthetics models and to expand the pool of participants.

ACKNOWLEDGEMENTS

The authors are thankful to the amateur photographer Mark Dekker (http://gplus.to/markdekker), who donated most of the photographic material used in this study.

12. REFERENCES

[1] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo, "Aesthetics and emotions in images," *Sig. Proc. Magazine, IEEE,* vol. 28, 2011.

[2] J. A. Redi, "Visual quality beyond artifact visibility," in IS&T/SPIE Electronic Imaging, 2013, 86510N-86510N-11.

[3] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Process-Image.* , vol. 25, no. 7, pp. 469-481, 2010

[4] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in Computer Vision–ECCV 2008, ed: Springer, 2008, pp. 386-399

[5] M. Redi and B. Merialdo, "Where is the interestingness?: retrieving appealing VideoScenes by learning Flickr-based graded judgments," in Proc. ACM Multimedia, 2012

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in Computer Vision–ECCV 2006, pp. 288-301

[7] C. D. Cerosaletti, A. C. Loui, and A. C. Gallagher, "Investigating two features of aesthetic perception in consumer photographic images: clutter and center," SPIE Conference Series, 2011, p. 5

[8] P. Obrador, L. Schmidt-Hackenberg, and N. Oliver, "The role of image composition in image aesthetics," in Proc. IEEE ICIP, 2010

[9] A. L. Yarbus and L. A. Riggs, Eye movements and vision vol. 2: Plenum press New York, 1967

[10] C. Wallraven, D. Cunningham, J. Rigau, M. Feixas, and M. Sbert, "Aesthetic appraisal of art-from eye movements to computers," Computational aesthetics, pp. 137-144, 2009

[11] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," Signal Processing Magazine, IEEE, vol. 28, pp. 50-59, 2011

[12] H. Alers, H. Liu, J. Redi and I. Heynderickx, "Studying the risk of optimizing the image quality in saliency regions at the expense of background content", IS&T/SPIE HVEI, 2010

[13] L.-K. Wong and K.-L. Low, "Saliency-enhanced image aesthetics class prediction," In Proc. IEEE ICIP, 2009

[14] Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in Proc. ACM Multimedia, 2010

[15] C. Koch and S. Ullman, "Shifts in Selection in Visual Attention: Toward the Underlying Neural Circuitry," Human Neurobiology, vol. 4, no. 4, pp. 219-27, 1985

[16] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," Trans. Pattern Analysis and Machine Intelligence, vol. 20, 1998

[17] Engeldrum, Peter G., "Psychometric scaling: a toolkit for imaging systems development," Imcotek Press, 2000

[18] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," Geneva, (2002)

[19] B.Yao, X. Yang, and S. Zhu, "Introduction to a large scale general purpose ground truth dataset: methodology, annotation tool, and benchmarks." EMMCVPR, 2007

[20] Redi, J., and Heynderickx, I., "Image Quality And Visual Attention Interactions: Towards A More Reliable Analysis In The Saliency Space," Proceedings of QoMEX, (2011)

[21] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision Research*, vol. 45, no. 18, pp. 2397–2416, August 2005

[22] Ninassi, O. Le Meur, P. L. Callet, D. Barba, and A. Tirel, "Task impact on the visual attention in subjective image quality assessment," Proc. EUSIPCO-06, 2006

[23] C. Mantel, N. Guyader, P Ladret, G. Ionescu and T. Kunlin "Characterizing eye movements during temporal and global quality assessment of h.264 compressed video sequences," in Proc. SPIE 8291, HVEI XVII, 2012