Image Integrity and Aesthetics: towards a more encompassing definition of Visual Quality

Judith A. Redi*[a], Ingrid Heynderickx[a,b]
[a] Delft University of Technology, Mekelweg 4, Delft, The Netherlands 2628 CD;
[b] Philips Research Laboratories, Prof. Holstlaan 4, Eindhoven, The Netherlands 5656 AE

ABSTRACT

Visual quality is a multifaceted quantity that depends on multiple attributes of the image/video. According to Keelan's definition, artifactual attributes concern features of the image that when visible, are annoying and compromise the integrity of the image. Aesthetic attributes instead depend on the observer's personal taste. Both types of attributes have been studied in the literature in relation to visual quality, but never in conjunction with each other. In this paper we perform a psychometric experiment to investigate how artifactual and aesthetic attributes interact, and how they affect the viewing behavior. In particular, we studied to what extent the appearance of artifacts impacts the aesthetic quality of images. Our results indicate that indeed image integrity somehow influences the aesthetic quality scores. By means of an eye-tracker, we also recorded and analyzed the viewing behavior of our participants while scoring aesthetic quality. Results reveal that, when scoring aesthetic quality, viewing behavior significantly departs from the natural free looking, as well as from the viewing behavior observed for integrity scoring.

Keywords: Visual quality, subjective image quality assessment, aesthetic quality, eye-tracking, visual attention

1. INTRODUCTION

Often visual quality is referred to as an overall concept that concerns the degree to which the visual experience satisfies the user of an imaging system. Keelan [1] gives an extended definition of visual quality and lists four families of attributes that contribute to the final visual appreciation. Artifactual (e.g. noise or blockiness) and preferential attributes (e.g. color saturation or brightness) are closely related to the performance and limitations of imaging technologies. Aesthetic (e.g., symmetry or harmony) and personal attributes (e.g. engagement) are instead more user and content-dependent. This extended and encompassing definition of visual quality, however, is not consistently used in the literature. The electronic imaging community mainly limits visual quality to the perceived integrity of the media (i.e., an image or video), when affected by visual degradations due to signal errors or technological limitations [2, 3]. Research in this field focused on quantifying the impact of visual impairments on perceived quality to be used for the optimization of media rendering. For the sake of clarity, we will here refer to this technology-oriented definition of visual quality as *image integrity*. From another perspective, the media management community defines visual quality as related to content pleasantness, mainly disregarding its integrity [4]. In particular, great work has been done in quantifying and predicting what we will call here the *aesthetic quality* of media, for applications such as image retrieval or intelligent thumbnailing [5-7].

Future multimedia applications will be mainly based on internet distribution and on on-demand content selection. In order to guarantee a high quality of experience in this application context, an encompassing definition of visual quality needs to integrate contributions of the four types of attributes proposed by Keelan, and consider their joint impact. Some (unconscious) efforts in this direction have already been done. In literature on computational aesthetics for example, preferential attributes such as color saturation are often considered as contributing to the final aesthetic quality judgment [5, 6]. From the opposite perspective, personal attributes such as engagement with the content have been found to influence the perceived integrity of the media [8]. Starting from these results, more systematic studies need to be performed to understand the interactions between different types of attributes that ultimately should result in a more encompassing model of visual quality appreciation.

In this paper, we investigate the nature of the interactions among different visual quality attributes, and in particular we start with the interaction between the aesthetic and artifactual attributes. We are specifically interested in (1)

Human Vision and Electronic Imaging XVII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8291, 829115 • © 2012 SPIE-IS&T
CCC code: 0277-786X/12/$18 • doi: 10.1117/12.911894

understanding the role of artifactual attributes in aesthetic quality evaluation, and (2) investigating differences in viewing behaviour, if any, between free image observation, image integrity evaluation and aesthetic quality assessment. To do so, we designed a psychometric experiment involving two sets of images, one including high quality images taken from the delft eye-tracking databases [9, 10], and the second including distorted versions of those same images, for which the integrity values were already known. Two groups of observers - one group per set - assessed the aesthetic qualityof the images. Differences in aesthetic quality were correlated with differences in perceived integrity. Furthermore, eye movements of the observers were recorded during the assessment task. By comparing saliency maps obtained during aesthetic scoring to existing maps recorded for the same image under free looking and integrity scoring [9, 10], we could analyse differences in attention mechanisms depending on the task.

In the following we report the details on the experiment and on its outcomes. In particular, section 2 reports on the experimental setup and methodology. In section 3 we investigate on the effect that a decrease in integrity has on the aesthetic appearance of images. In section 4 we analyse the viewing behaviour of observers scoring the aesthtetic quality of images. Finally, we discuss the outcomes of the experiment in section 5, where we also sketch possible follow-up research lines for this work.

2. A PSYCHOMETRIC EXPERIMENT ON THE AESTHETIC QUALITY OF IMAGES

2.1 Image material

To investigate the role played by image integrity on the aesthetic appeal of images, we designed a between-subjects experiment, involving one set of high-quality images and a second set including distorted versions of those same images.

We selected 57 original images, of which 29 were taken from the Delft Eye-tracking database I [9] (shortened as DET1 hereafter, and corresponding to the original images of the well-known LIVE database [11]), while the remaining 28 images were taken from the Delft Eye-tracking Database II (shortened as DET2 hereafter) [10]. Samples of these images are shown in Figure 1. All images had a high integrity level and were included in the first set of images, which we will refer to as the High Integrity (HI) image set.

We then produced a second collection of images by encoding the images in set HI with the JPEG compression standard. Encoding bitrates were selected among those used in the databases [9,10] so that the resulting images spanned a

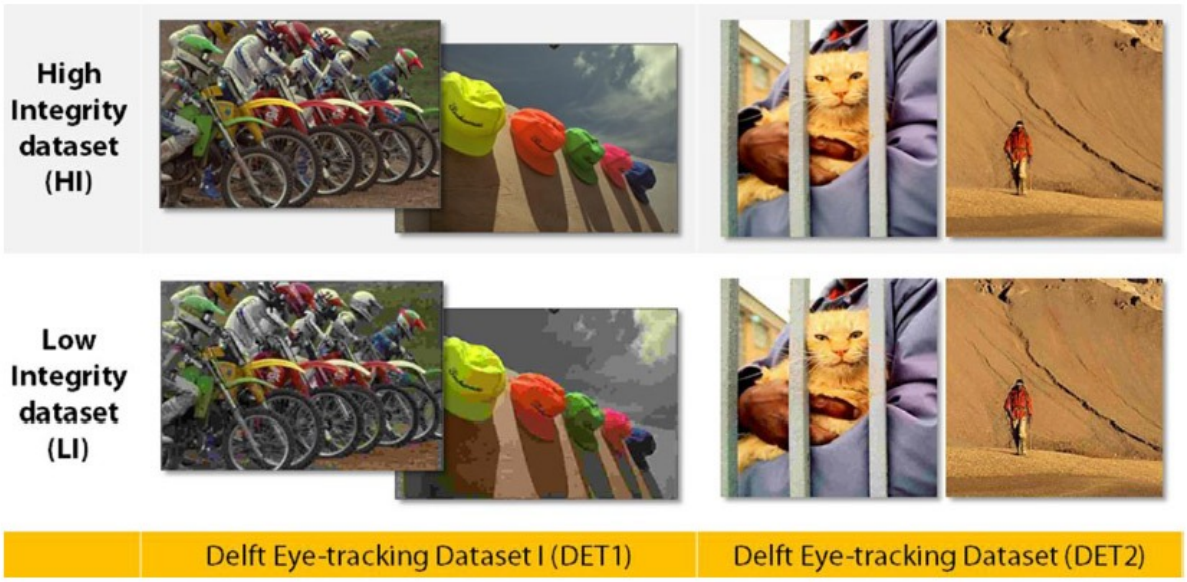

Figure 1. Examples of the images involved in the psychometric experiment. The High Integrity dataset included 57 undistorted images, and was evaluated by a first group of 12 observers. A second group of 12 observers evaluated the Low integrity dataset, including the same 57 image contents as HI but compressed to various integrity levels

relatively large range of integrity, and so that blocking artifacts were at least noticeable (see Figure 1). In particular, images taken from DET1 were divided into six groups (five groups of five contents and one of four contents), and each group was encoded at a different level, ranging from Q = 5 to Q = 40 [9, 12]. Images taken from DET2 were instead compressed all at different levels with $Q \in [11, 78]$, as per [10, 13]. All distorted images were included in a second set, to which we will refer in the following as the Low Integrity (LI) image set. In total, (57x2=)114 images were evaluated in aesthetic quality during the experiment.

2.2 Instrumentation and experimental setup

Stimuli were displayed on a 17" CRT monitor having a resolution of 1024x768 pixels. To track eye-movements we used a *SensoMotoric Instruments* GmbH Eye Tracker with a sampling rate of 50/60 Hz, a pupil tracking resolution of 0.1°, a gaze position accuracy of 0.5 - 1°, and an operating distance between the subject and the camera of 0.4 - 0.8 meters. Participants were constrained by a chinrest at a distance of 0.7 meters from the display. The illumination was kept constant at approximately 70 lux, following the setup already used in [9, 10], which was proven to guarantee stability of the eye-tracker. The user interface for the subjective test was implemented using the *Neurobehavioral Systems* software Presentation.

2.3 Methodology

Two disjoint groups of 12 participants each took part in the experiment. They were mostly recruited from the Delft University of Technology and aged between 22 and 40 years. Group 1 evaluated the HI image set, and group 2 the LI image set. Participants were asked to assess the *aesthetic appeal* of images, using the Single Stimulus method with a continuous numerical scale [14], ranging from 0 to 10 with "0" being very low aesthetic appeal and "10" very high aesthetic appeal. The participants were informed that they might encounter low integrity images in their assessment task, and they were explicitly asked *not* to take into account the integrity loss in their aesthetic judgment. The same instructions were given to both groups of participants (also to those scoring the HI images). The scoring task was performed on a single session, involving all 57 images included in the image set. Throughout the whole experiment, which took, on average, 30 minutes per subject, the movements of the subjects' eyes were recorded through an eye-tracker.

The following protocol was applicable to every participant. After a brief oral introduction, the eye-tracker was calibrated on the participant's gaze based on a 13-points grid. Subsequently, the participant went through a short training (i.e. four images) to get acquainted with the task and scoring interface. Participants had no time constraints in observing the images prior to scoring. When ready, they could access the scoring screen, existing of the aesthetic appeal scale only. This scoring screen was kept separated from the image to avoid distraction during the image observation. Participants entered their judgment by simply positioning a slider on the scoring scale through a computer mouse. After the score was entered, the interface showed the next image. Images appeared in a randomized order, different for each participant, to minimize fatigue and learning effects.

3. AESTHETICS AND INTEGRITY: ANALYSIS OF THE MEAN OPINION SCORES

The experiment resulted in two collections of aesthetic scores, one for the HI image set and one for the LI image set. To study the role of integrity in aesthetic quality perception, we compared these two collections of scores. If aesthetics and integrity were uncorrelated, we would expect to find no difference in aesthetic scores between HI and LI images. Conversely, differences in the scores would indicate an effect of the diminished integrity on the overall aesthetic appearance of the image.

3.1 Aesthetic data

Aesthetic quality scores were averaged across observers for each image. This resulted into a Mean Aesthetic Opinion Score (MAOS) per image:

$$MAOS^I = \frac{\sum_l AS_o^I}{L}, \qquad l = 1,...,L, \tag{1}$$

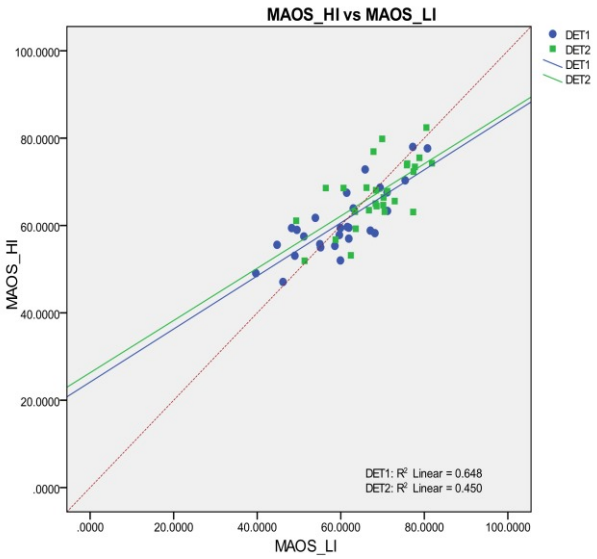

Figure 2. Mean Opinion Aesthetic scores obtained for the Low Integrity dataset plotted against MAOS obtained for the High Integrity Dataset.

where I is an image $\in$ {HI, LI}, AS_o^I is the Aesthetic Score given by observer O to image I, and L is the number of participants (L = 12). Neither outlier observers were detected, nor outlier scores removed. For comparison purposes, MAOS were then re-mapped into the range [0, 100].

A first, interesting outcome concerns the inter-observer agreement in judging aesthetic quality. We computed for each image *I* the standard deviation of the aesthetic scores across all observers, and took the average value across each dataset. As such, we obtained a value of 15.01 for the LI dataset and of 18.8 for the HI dataset, both on a 100-point scale. These values are only slightly higher than values obtained for the integrity scores, i.e. an average standard deviation of 13.6 for the scores in [15] and of 14.05 for the scores in [13], again on a 100-point scale. Thus, we can assume that aesthetic quality is just slightly less precisely quantifiable than integrity. Furthermore, we know from literature that the Single Stimulus methodology is prone to high observer disagreement in integrity scoring tasks [16, 17]. Perhaps adopting a methodology less prone to observer disagreement (e.g., a paired comparison test), we might be able to quantify MAOS with a higher confidence.

Figure 2 shows a scatter plot of the scores obtained for the LI dataset against those obtained for the HI dataset. Each point in the graph represents one image *I*. Its y-coordinate corresponds to the MAOS obtained for the HI version of *I*, i.e. $MAOS_I^{HI}$, while the x-coordinate represents the MAOS obtained for the LI version of *I*, i.e. $MAOS_I^{LI}$. For the sake of completeness, the plot differentiates between the images taken from the Delft Eye-Tracking database I (DET1) and those taken from the Delft Eye-Tracking database II (DET2). For both subsets, the correlation between the $MAOS_I^{HI}$ and $MAOS_I^{LI}$ is quite high (0.81 for DET1, 0.67 for DET2). Systematic differences between MAOS are not found, i.e., $MAOS^{LI}$ are not systematically higher or lower than $MAOS^{HI}$, according to a paired-samples t-test (p = 0.704 for DET1 and p = 0.292 for DET2). However, the linear models approximating these distributions significantly depart from the bisector of the plane (p < 0.001 in both cases). In particular, it seems that at low aesthetic quality values, the HI images were scored higher than the LI images. Conversely, at higher aesthetic quality values, the HI images were scored lower than the LI images.

3.2 Integrity data

Mean Integrity Opinion Scores (MIOS) were available from previous experiments [12, 13] for all images in the LI dataset. For the images taken from Delft Eye-Tracking dataset I, integrity scores were obtained from 18 observers using a single-stimulus 5-point categorical scoring scale [14]. Observers' opinions were first transformed into numerical values

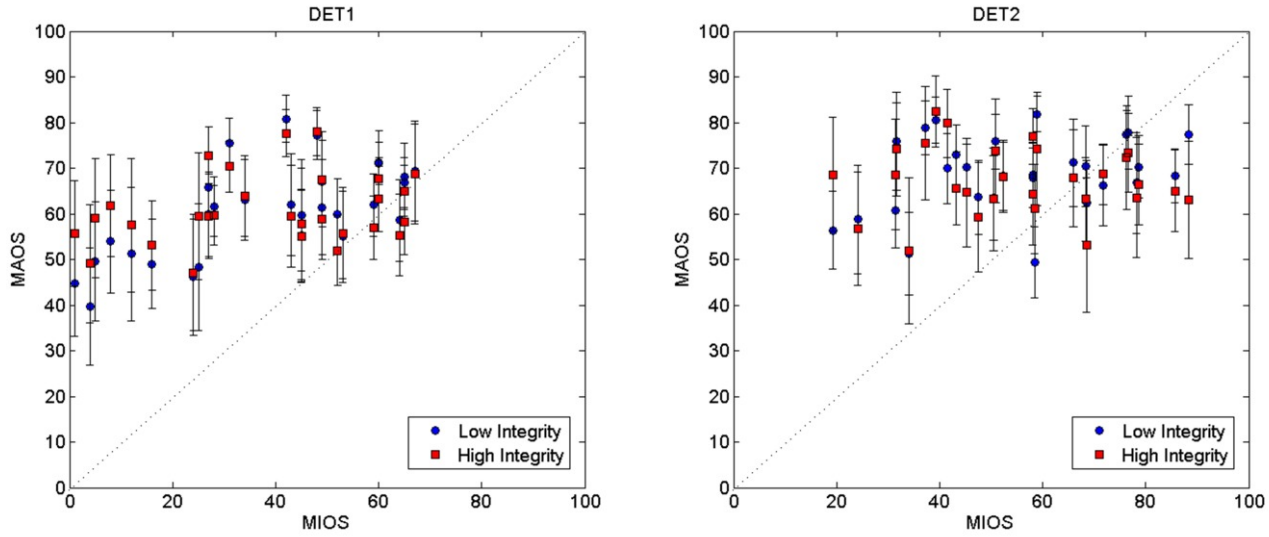

Figure 3. Mean Opinion Aesthetic Scores obtained for Images in the Low Integrity dataset (Blue markers) and in the High Integrity Dataset (Red Markers). For each image, the MAOS is plotted against the MIOS of the corresponding image in the LI set.

(5 corresponding to "Excellent", 1 corresponding to "Bad") and then averaged into MIOS according to (1). MIOS were finally re-mapped into the range [0,100] to allow comparison with the other data considered in this analysis. Images from the DET2 database were instead assessed in integrity using a Single Stimulus continuous numerical scale [14]. Integrity scores were collected from 20 participants and averaged into MIOS $\in$ [0,10]. Also in this case, we re-mapped the original MIOS into the range [0,100] for comparison purposes.

It should be noticed that, although re-mapped into the same numerical range, MIOS from DET1 and DET2 are not directly comparable and cannot be merged on a single scale. In other words, an image having a MIOS = 60 in DET1 does not necessarily have the same quality level as an image having a MIOS = 60 in DET2. There are two reasons for this: (1) the integrity scores were collected with two different methodologies (categorical vs continuous numerical scale) and (2) in general, scores obtained from a single stimulus scoring setup are known to be adjusted to the quality range represented in the image dataset [16, 17]. Since DET1 and DET2 spanned two different quality ranges, we expect the corresponding MIOS to be not directly comparable. For this reason, the rest of the analysis is performed for the DET1 and DET2 data separated.

3.3 Relationship between image integrity and aesthetic quality

Figure 3 gives an overview of the MAOS obtained in the experiment and their relationship with integrity. For each image *I* in the dataset, we plot both $MAOS_I^{HI}$ (blue dots) and $MAOS_I^{LI}$ (red squares) against the $MIOS_I^{LI}$ (Note that $MIOS_I^{HI}$ are not available from previous experiments, and will not be considered throughout this study). For DET1, the MAOS obtained for the LI images are correlated with the MIOS of those images (Pearson Correlation of 0.66), while this is not true for the HI images (Pearson correlation of $MAOS_I^{HI}$ with $MIOS_I^{LI}$ is 0.27). For DET2, lower correlations are found between $MIOS^{LI}$ and MAOS, for both LI and HI images (Pearson correlation of 0.30 for LI and -0.83 for HI, respectively). As can be reasonably expected, the MAOS of HI images are not correlated to the integrity values of the LI images. Conversely, we find some correlation between the aesthetic values and the integrity values of low integrity images. This might indicate that the quality level of the image influences to some extent its aesthetic quality.

To further investigate this aspect we study whether the difference in aesthetic quality between HI and LI images is somehow related to the integrity of the LI images. Thus, we first compute these differences for each image *I* as:

$$MAOS\_diff_I = MAOS_I^{HI} - MAOS_I^{LI}$$

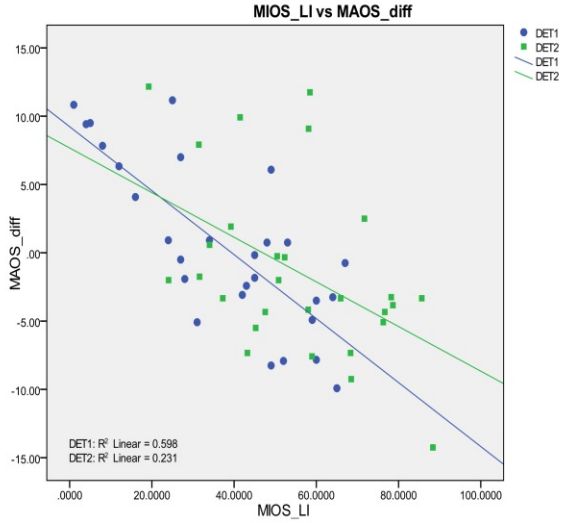

Figure 4. Differences in Mean Opinion Aesthetic Scores obtained for Images in the Low Integrity dataset and in the High Integrity Dataset. For each image, the MAOS difference is plotted against the MIOS of the corresponding image in the LI set.

and then plot each of them against the integrity of their LI image, i.e. $MIOS_I^{LI}$ (see figure 4). MAOS differences appear to be (negatively) correlated with the $MIOS_I^{LI}$ with coefficients of 0.76 for DET1 and 0.48 for DET2. In practical terms, the results indicate that:

- if the integrity of an LI image is very low, its aesthetic quality is scored lower than that of the corresponding HI image;
- if the integrity of an LI image is in the mid-scale range, its aesthetic quality is scored approximately equal to that of the corresponding HI image;
- if the integrity of an LI image is relatively high, its aesthetic quality is scored higher than that of the corresponding HI image.

This behavior can be explained as some sort of "Halo effect" [18] of integrity on aesthetics. Observers that scored the LI dataset viewed a whole range of different integrity values throughout the experiment and became aware of the integrity of the images. Although explicitly asked not to take it into account, they modulated their aesthetic quality judgment with integrity to some extent: low integrity images were "punished" with lower aesthetic scores, while high integrity images were "rewarded" with higher aesthetic scores.

4. AESTHETICS AND INTEGRITY: ANALYSIS OF THE VIEWING BEHAVIOR

In this section we analyse the eye-tracking data collected during scoring the aesthetics of HI and LI images. More specifically, we want to find out (1) whether the integrity level of images influences the viewing behaviour, when scoring aesthetic quality and (2) whether the viewing behaviour when scoring aesthetic quality significantly departs from that found during free looking or integrity scoring.

4.1 Saliency data for aesthetic scoring

Eye-trackers typically produce a collection of pupil movements, summarized in terms of fixations and saccades. We follow the procedure formalized in [19] to process these data and to transform them into saliency maps. A saliency map is a probability map having the same size as the image it refers to; each pixel represents the probability that the corresponding pixel in the actual image is attended by an (average) observer. In producing the saliency maps we don't take into account temporal information, i.e. all fixations are considered equally important in creating the map, independent of their duration. The procedure that we applied to each image $I \in \{HI, LI\}$ to produce a saliency map $SS^{(I)}$ can be summarized as (for a detailed description, please refer to [19]):

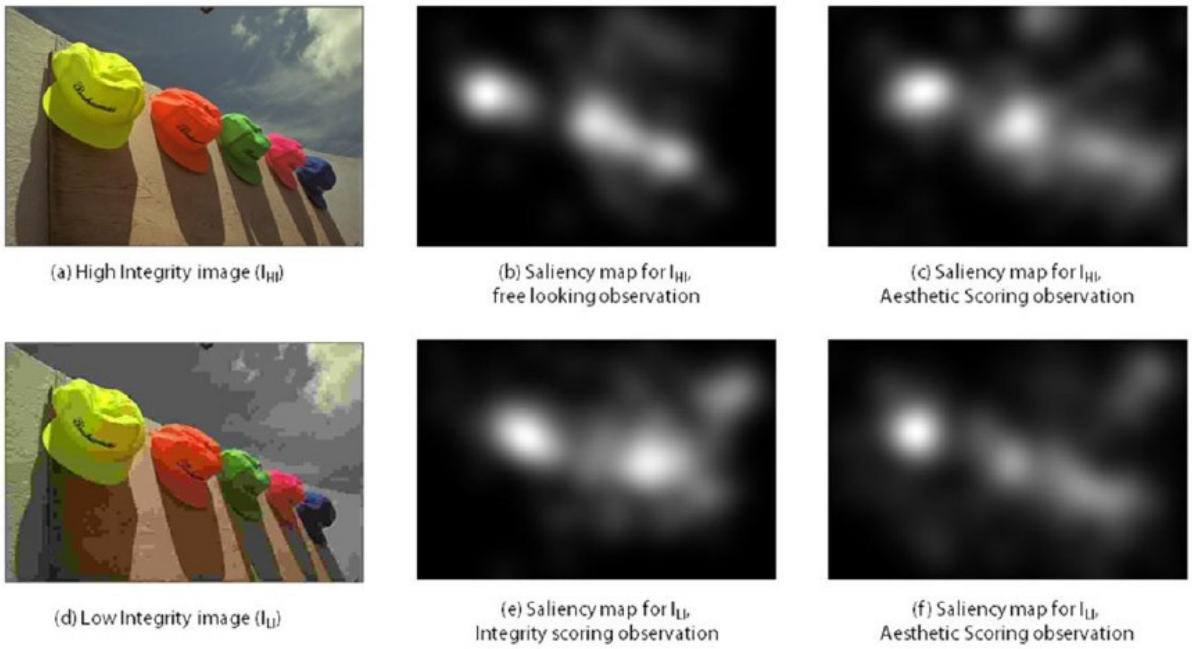

Figure 5. examples of saliency maps obtained for different tasks and integrity levels

1. Fixation points were extracted for each observer separately, and then added to an overall fixation map $FM^{(I)}(x,y)$; eventually this map included all fixation points from all observers
2. For each fixation point in $FM^{(I)}(x,y)$, a grey scale patch was applied having a Gaussian intensity distribution; the variance, σ, of the intensity distribution approximated the size of the fovea (about 2° of visual angle). Thus, the saliency value $SS^{(I)}(k,l)$, at location (k,l) of the saliency map corresponding to image I resulted in:

$$SS^{(I)}(k,l) = \sum_{j=1}^{T} \exp[-\frac{(x_j - k)^2 + (y_j - l)^2}{\sigma^2}] \quad (1)$$

where (x_j, y_j) are the spatial coordinates of the jth fixation ($j=1...T$) in the map $FM^{(I)}(x,y)$, and $k \in [1, W_I]$, $l \in [1, H_I]$, assuming I to be of size W_I x H_I (and consequently also $SS^{(I)}$ and $FM^{(I)}$ have the same size).

This procedure resulted into 114 maps, 57 for the images in HI and 57 for those in LI.

During image observation, attention can be attracted by different features of the image, depending on experimental conditions (e.g., depending on task or integrity level [15]). As a consequence, saliency may vary with the variation of the experimental setup. Dissimilarities in saliency maps obtained for the same image but under different viewing condition can thus indicate an effect of the changed condition on viewing behaviour [13, 15, 19]. Multiple methods have been proposed in literature to quantify the similarity of saliency maps [19]. For the sake of brevity, in this paper we limit the analysis to the computation of the Linear Correlation Coefficient (LCC) between pairs of maps. LCC values range between [-1, 1], 1 being maximum linear correlation, 0 absence of correlation and -1 inverse correlation. Values of LCC close to 1 indicate high similarity between two saliency maps.

To answer our first question and evaluate the effect that a decrease in integrity has on viewing behavior when scoring aesthetic quality, we compare the maps obtained for the HI images, i.e. $SS^{(I, HI)}$, with the corresponding maps for the LI images, i.e. $SS^{(I, LI)}$. The values LCC($SS^{(I, HI)}$, $SS^{(I, LI)}$) are summarized in figure 6, and plotted against the integrity level of I_{LI}. Some correlation is found between the LCC and the integrity of the LI images for the DET1 database (Pearson correlation coefficient of 0.58). In particular, the similarity between $SS^{(I, HI)}$ and $SS^{(I, LI)}$ seems to decrease with a decrease in integrity. In other words, the lower the integrity of an image, the more different people look at I_{LI} and I_{HI} when scoring their aesthetic quality. This might be due to a "distraction of attention" caused by local blocking artifacts. This sort of

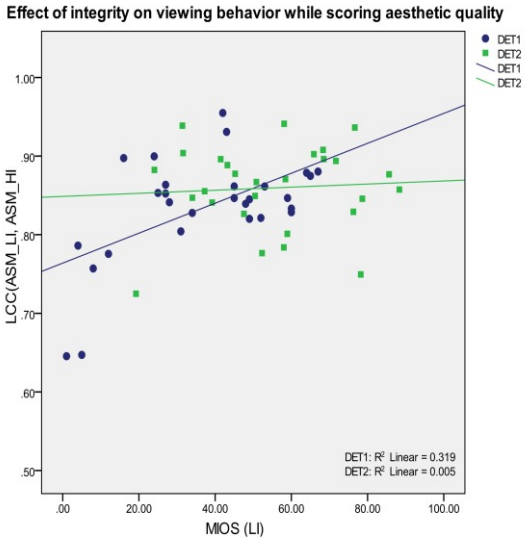

Figure 6. Similarity in viewing behavior while scoring the aesthetic quality of high and low integrity images. LCC is computed among pairs of maps obtained for the aesthetic scoring of the same image but at different integrity levels (HI and LI). Similarity values are then plotted against the integrity value of the LI image.

behavior is visible in figure 5(c) and 5(f). The cloud on the top right corner is heavily distorted in I_{LI}, and not in I_{HI} (compare figure 5(d) with 5(a)). The corresponding saliency maps when scoring aesthetics (figure 5(f) and (c), respectively) show that the cloud receives more attention in I_{LI} than in I_{HI}. This behavior is not found for images in DET2. This might be due to the fact that artifacts in DET2 are less strong (images are compressed at higher qualities in DET2 than in DET1) and thus do not attract attention.

4.2 Saliency data for integrity scoring and free looking.

For both the HI and LI dataset, saliency data for free looking and integrity scoring were available. All data (both from DET1 and DET2) were collected using the same eye-tracker as used in this study, and using an almost identical setup [9, 10, 12, 13]. It should be mentioned that for DET1 participants observed the images for a constrained amount of time (6 s) while for both DET2 and the present experiment viewing time was unconstrained.

In this analysis, we want to determine the effect on saliency of (1) the viewing task (e.g., aesthetic scoring versus free looking) and (2) a decrease in integrity (i.e. HI versus LI for the same task). Thus, of the data provided in DET1 and DET2, we use:

- The saliency maps obtained from observers freely looking at HI images (HIFL). These data represent the saliency distribution of images viewed under natural conditions, i.e. without task constraints. These data are often taken as reference condition, when investigating the influence of task or integrity on visual attention [12, 13, 15].
- The saliency maps obtained from observers scoring the integrity of LI images (LIIS), which are known to depart

Table 1. Saliency data involved in the analysis.

Dataset name	Integrity level of the images	task	Saliency maps available for	
			DET1	DET2
HIFL	High	Free Looking	x [9]	x [10]
LIIS	Low	Integrity Scoring	x [9]	x [10]
HIAS	High	Aesthetic Scoring	x	x
LIAS	Low	Aesthetic Scoring	x	x

significantly from those obtained while free looking at HI images, due to both task and integrity effects.

A summary of the types of saliency data used in the following analysis is given in table I.

4.3 Aesthetic scoring and viewing behaviour

For each of the possible pairs of datasets in table I, we computed the LCC between maps corresponding to the same image and obtained under different experimental conditions. To compute the similarity among HIFL maps and HIAS maps, for example, we computed the LCC($SS^{(I,\ \mathrm{HIFL})}$, $SS^{(I,\ \mathrm{HIAS})}$) for each image $I \in$ HI and then averaged these values over the whole dataset. Figure 7 reports these average LCC values for all possible pairs of datasets. The results should be compared to the dashed line in figure 7, representing the Upper Empirical Similarity Limit (UESL, [19]), and indicating the inter-observer variability. The UESL is defined here as the level of similarity in viewing behaviour among the observers that took part in the HIFL experiment. The UESL is computed by (i) randomly splitting the 18 observers of DET1 into two equally sized groups, (ii) computing the HIFL saliency maps per group, (iii) calculating the LCC between the two resulting maps and (iv) averaging the LCC values over the whole image dataset. To ensure a robust estimation of the UESL, this operation was repeated 50 times with every time a different composition of the two groups. The UESL reported in figure 7 is obtained after averaging the LCC values of the 50 trials.

From figure 7 we can fairly conclude that the aesthetic scoring task influences viewing behaviour. The similarity between HIFL maps and both HIAS and LIAS maps is lower than the UESL, thus we can conclude that viewing behaviour changes more between free looking and scoring aesthetic quality than what can be expected from inter-observer variability. In both cases, this difference is larger than that found between free looking at HI images and integrity scoring of LI images. Furthermore, viewing behaviour, while scoring aesthetic quality, differs also from viewing behaviour while scoring integrity, at least for the LI images. Finally, also integrity has some effect on the viewing behaviour, at least for DET1 images, which confirms the results found in section 4.1.

5. CONCLUSIONS

In this paper we investigated the impact of image integrity on aesthetic quality. We asked a group of observer to evaluate the aesthetic quality of a set of high integrity images, and a second group to evaluate the aesthetic quality of the same images, but with lower integrity. By comparing the aesthetic scores obtained from the two groups of observers, we found that image integrity influences aesthetic quality with some sort of "halo effect" [18]. Although explicitly asked not to take integrity into account, observers that viewed low integrity images scored their aesthetic quality lower (with respect to their high integrity counterpart) for highly distorted images and higher for slightly distorted images. This indicates that

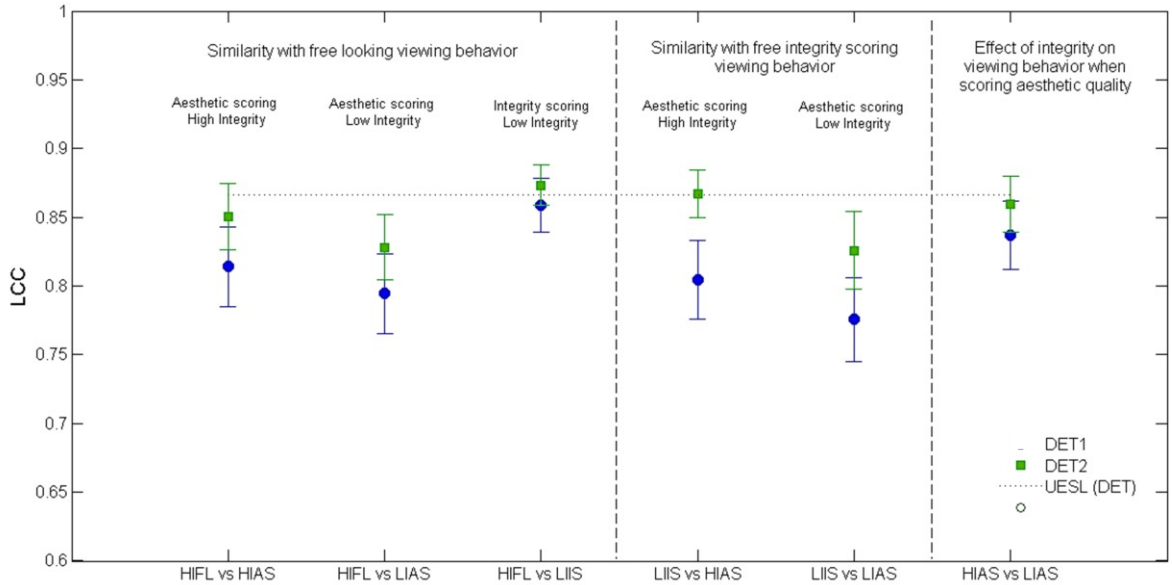

Figure 7. LCC between saliency maps obtained under different experimental conditions. Maps recorded under aesthetic scoring significantly depart from other conditions, i.e. free looking or integrity scoring.

integrity plays a role in the aesthetic appeal of images, and should be taken into account when designing e.g. mechanisms for image retrieval based on aesthetics.

A second aspect we investigated was the impact that aesthetic scoring has on viewing behavior. Our data showed that viewing behavior during aesthetic scoring significantly departed from that observed from free looking and integrity scoring of the same images. This difference was found to be more pronounced for low integrity images than for high integrity images. Further work in this direction will clarify *how* viewing behavior changes, i.e., which areas of the image attract more attention while scoring aesthetic quality. Outcomes of such research would be relevant in e.g. intelligent thumbnailing, to perform image miniaturization in a way that preserves those parts of the image that mostly matter for its aesthetic appeal.

REFERENCES

[1] Keelan, B., "Handbook of image quality: characterization and prediction," Marcel Dekker, Inc., New York, 2002.

[2] Lin, W., and Jay Kuo, C.-C., "Perceptual Visual Quality Metrics: A Survey," Journal of Visual Communication and Image Representation, vol. 22, no. 4, pp. 297-312, (2011).

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

[4] Datta, R., Joshi, D., Li, J., and Wang, J.Z., "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, (2008).

[5] Datta, R., Joshi, D., Li, J., and Wang, J.Z., "Studying Aesthetics in Photographic Images Using a Computational Approach," Lecture Notes in Computer Science, vol. 3953, Proceedings of the European Conference on Computer Vision, Part III, pp. 288-301, Graz, Austria, (2006).

[6] Luo, Y., and Tang, X., "Photo and Video Quality Evaluation: Focusing on the Subject", Proceedings of the 10th European Conference on Computer Vision: Part III, (2008)

[7] Jiang, W., Loui, A.C., and Cerosaletti, C.D., "Automatic aesthetic value assessment in photographic images", in proc. IEEE International Conference on Multimedia and Expo, (2010)

[8] Kortum, P., and Sullivan, M., "The Effect of Content Desirability on Subjective Video Quality Ratings". Human Factors: The Journal of the Human Factors and Ergonomics Society, (2010)

[9] Liu H., and Heynderickx, I., "TUD Image Quality Database: Eye-Tracking Release 1", http://mmi.tudelft.nl/iqlab/eye_tracking_1.html, (2010)

[10] Alers, H., Liu, H., Redi, J., and Heynderickx, I., "TUD Image Quality Database: Eye-Tracking Release 2", http://mmi.tudelft.nl/iqlab/eye_tracking_2.html, (2010)

[11] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Processing. vol. 15, no. 11, (2006)

[12] Liu, H., and Heynderickx, I., "Visual Attention in Objective Image Quality Assessment: based on Eye Tracking Data," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, pp. 971-982, July, (2011)

[13] Alers, H., and Heynderickx, I., "Studying the risks of optimizing the image quality in saliency regions at the expense of background content". IS&T/SPIE Electronic Imaging 2010 and Image Quality and System Performance VII, (2010)

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

[15] Redi, J.A., Liu, H., Zunino, R. and Heynderickx, I., "Interactions of visual attention and quality perception", In: IS&T/SPIE Electronic Imaging 2011 and Human Vision and Electronic Imaging XVI. Vol 7865. (2011)

[16] De Ridder, H., "Cognitive Issues in image quality measurement", J Electronic Imaging, 10(1), 47-55 (2001)

[17] Redi, J., Liu, H. ,Alers, H., Zunino, R. and Heynderickx, I., "Comparing subjective image quality measurement methods for the creation of public databases", in Proc. IS&T/SPIE Electronic Imaging, vol. 7529 (2010)

[18] Dion, K., Berscheid, E., & Walster E., "What is Beautiful Is Good", Journal of Personality and Social Pyschology, 24 (3), 285-290, (1972)

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