Interactions of visual attention and quality perception

Judith Redi*[a], Hantao Liu[a], Rodolfo Zunino[b], and Ingrid Heynderickx[a,c]
[a]Delft University of Technology, Mekelweg 4, Delft, The Netherlands 2628 CD
[b]University of Genoa, DIBE, Via Opera Pia 11a, Genova, Italy 16145
[c]Philips Research Laboratories, Prof. Holstlaan 4, Eindhoven, The Netherlands 5656 AA

ABSTRACT

Several attempts to integrate visual saliency information in quality metrics are described in literature, albeit with contradictory results. The way saliency is integrated in quality metrics should reflect the mechanisms underlying the interaction between image quality assessment and visual attention. This interaction is actually two-fold: (1) image distortions can attract attention away from the Natural Scene Saliency (NSS), and (2) the quality assessment task in itself can affect the way people look at an image. A subjective study was performed to analyze the deviation in attention from NSS as a consequence of being asked to assess the quality of distorted images, and, in particular, whether, and if so how, this deviation depended on the distortion kind and/or amount. Saliency maps were derived from eye-tracking data obtained during scoring distorted images, and they were compared to the corresponding NSS, derived from eye-tracking data obtained during freely looking at high quality images. The study revealed some structural differences between the NSS maps and the ones obtained during quality assessment of the distorted images. These differences were related to the quality level of the images; the lower the quality, the higher the deviation from the NSS was. The main change was identified as a shrinking of the region of interest, being most evident at low quality. No evident role for the kind of distortion in the change in saliency was found. Especially at low quality, the quality assessment task seemed to prevail on the natural attention, forcing it to deviate in order to better evaluate the impact of artifacts.

Keywords: Visual attention, image quality assessment, eye-tracking, scene saliency, subjective quality evaluation

1. INTRODUCTION

When observing a scene, the human eye typically optimizes the information acquisition process to overcome the brain's limited capacity of information processing. As a result, when humans observe images they perform some sort of "scanning", focusing on selected (salient) regions and neglecting poorly informative areas [1-3]. This high-level feature of the Human Visual System (HVS) has lately become of interest for researchers committed in image quality perception modeling. It is reasonable to assume that in a freely looking situation the eye's sensitivity to visual distortions depends on the saliency of the area where distortions are located. Therefore, when modeling quality perception in a freely looking situation, one would expect that artifacts appearing in less salient regions are less visible, thus less annoying, than artifacts affecting the Region of Interest (ROI). In addition, when people are requested to score the quality of distorted images, they may pay proportionally more attention to the most degraded areas in an image, since these areas determine their quality score. Provided the soundness of these assumptions, integration of saliency information into objective image quality metrics (e.g. [4]) might bring benefits in accuracy and possibly in computational cost saving.

Existing studies [5-9] seem to encourage the incorporation of saliency information into objective image quality metrics (OIQM). The core problem in defining an effective strategy for saliency integration in OIQM is to find coherence with human perception. OIQM typically compute a measure for the quality degradation of a distortion on a local basis. Therefore, a common choice is to weight the metric value with the corresponding measured saliency location-wise (e.g. pixel by pixel). Studies using this approach produce contradictory results. In [5], Liu and Heynderickx prove that weighting PSNR and SSIM metrics with Natural Scene Saliency (NSS, i.e. the saliency distribution over the image obtained during free looking) provides benefits in quality prediction accuracy. On the other hand, Ninassi et Al. [6] do not obtain any useful contribution in accuracy for the same metrics, tested on a different dataset. Larson and others [7] use a similar integration strategy to compare the use of NSS and of saliency measured during a quality scoring task. They

conclude that the use of NSS brings higher accuracy in the quality estimates, based on results on a subset of the LIVE [10] dataset. Conversely, Redi et al. [8] show how integrating pure NSS information in Machine Learning based metrics is counter-productive. Clearly all these results are not in line with each other. Hence, a deeper understanding of how visual attention and quality perception interact might be beneficial for a more correct design of the integration strategy, which might depend on e.g. the distortion affecting the picture, and/or on its quality level.

The assumption that annoying artifacts located outside the ROI are to a large extent neglected in a quality judgment is reasonable when observers freely look at images. On the other hand, OIQM target the reproduction of perceived quality scores, which can only be obtained during an evaluation task. Recent experiments already showed how eye-tracking data collected during image quality scoring differ from those collected during free looking, and how this attention deviation might depend on the specific kind of distortion [5, 11, 12]. Two hypotheses can be formulated for motivating this phenomenon, namely (1) the tendency of observers to scan the whole image in order not to disregard details in expressing their judgments, and (2) the distractive power of artifacts, that can move the focus of attention to heavier distorted areas. For particular kinds of distortions and contents, indeed, the annoyance of background artifacts can become so relevant that visual attention is actually deviated from the ROI [8]. Furthermore, observers might be more careful in equally scanning every part of the image when evaluating quality, in order to provide a sound judgment. To design an accurate model of image quality perception, not only the effects of saliency on quality perception should be considered, but also the influence of the (quality scoring) task on how the image is looked at.

A subjective study, based on tracking eye-movements during quality assessment, is performed to better understand the mechanisms underlying the interaction between image quality evaluation and visual attention. The aim of the experiment is to analyze whether quality assessment yields a deviation in saliency from NSS, and, if so, how this deviation depends on the distortion kind and/or amount. Eye-tracking data are processed to obtain saliency maps, averaged across observers, and these maps are compared to NSS maps derived from previous experiments [5]. Results show the existence of a deviation from the NSS as a consequence of scoring and its dependence on the quality level of the images. The outcome of this study is meant to provide empirical evidence of the impact of distortions on visual attention; indeed, they represent a starting point for designing a strategy to include spatial saliency into objective quality assessment metrics.

2. EYE-TRACKING DURING QUALITY SCORING: EXPERIMENTAL SETUP

For the experimental session, subjects were requested to score the quality of distorted images in a 6 (original images) x 3 (distortion types) x 3 (quality levels) within-subject design. During quality scoring, the eye-movements of the subjects were recorded. These eye-tracking data were then further processed to saliency maps, which were compared to the NSS, obtained during freely looking at the same six original images.

2.1 Image Material

The stimuli used in this experiment consisted of several distorted versions of 6 original images selected from the LIVE database. The original images used are shown in figure 1. They were selected such that we had a fair representation of different content, including images with and without a clear region of interest, images with and without humans in the picture, and images with various amounts of textured components. To evaluate the effect of distortion type and quality level, various distorted version of these original images were chosen from the LIVE database. Three kinds of distortions were considered, namely JPEG compression, White noise and Gaussian Blur. For each original image and distortion

Bikes *Light House* *Painted House* *Rapids* *Stream* *Woman Hat*

Figure 1. Images involved in the experiment. Contents and all their distorted versions are taken from the LIVE dataset [10]

Table I. Objective distortion parameters for the Images involved in the visual attention study. Contents and all their distorted versions were taken from LIVE dataset.

	Gaussian Blur		White Noise		JPEG Compression		
	Name	Kernel width	Name	Standard Deviation	Name	Compression Rate	
Bikes	img60	0.619759	img95	0.035156	img77	0.42552	High Quality
	img112	0.906218	img68	0.089844	img81	0.56447	Medium Quality
	img40	2.624972	img97	0.257813	img226	1.4875	Low Quality
Lighthouse	img4	0.447884	img40	0.0625	Img86	0.39	High Quality
	img102	0.82028	img43	0.019531	img44	0.42931	Medium Quality
	img97	1.479136	img96	0.171875	img231	1.2943	Low Quality
Painted House	img39	0.949187	img142	0.03125	img152	0.42863	High Quality
	img110	1.249969	img111	0.058594	img56	0.62819	Medium Quality
	img116	7.66665	img26	0.125	img228	1.4535	Low Quality
Rapids	img9	0.562467	img101	0.019531	img9	0.43058	High Quality
	img34	1.020802	img107	0.046875	img110	0.85997	Medium Quality
	img141	1.421844	img74	0.125	img146	1.4648	Low Quality
Stream	img71	0.419238	img88	0.0625	img16	0.57467	High Quality
	img126	0.834603	img2	0.1875	img185	1.0008	Medium Quality
	img58	3.083306	img106	0.3125	img137	1.6833	Low Quality
Woman Hat	img61	1.020802	img11	0.019531	img1	0.3263	High Quality
	img42	1.479136	img3	0.039063	img89	0.45313	Medium Quality
	img132	3.541641	img119	0.136719	img107	0.60102	Low Quality

type, three different quality levels were selected: a highly distorted one, one with a medium distortion level, and one for which the applied distortion just slightly compromised the quality. This resulted in 54 images in the final dataset, varying both in the distortion type and quality level (details are reported in table I).

2.2 Instrumentation and experimental setup

A *SensoMotoric Instruments* GmbH Eye Tracker was used for the experiment. It had 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. The various stimuli were displayed on a 17" CRT monitor with a resolution of 1024x768 pixels. To guarantee stability of the eye-tracking equipment a sufficiently high illumination level (~70 lux) was selected. The user interface for the subjective test was implemented using the *Neurobehavioral Systems* software Presentation. Subjects were kept at a fixed distance of 0.7 meters from the display using a chinrest. A total of 14 observers were recruited from the Delft University of Technology. Their age ranged between 22 and 35 years.

2.3 Methodology

After a brief oral introduction, each experiment started with a calibration of the eye-tracker. Participants were requested to focus on 13 different points spread over the monitor screen, and their eye fixations were recorded to calibrate the eye-tracking data. Subsequently, the participants got a short training to make them acquainted with the kind and amount of distortion they could expect in the test set. Users were asked to assess quality using the Single Stimulus method with continuous numerical scaling [13]. The scoring scale ranged from 0 to 10, where "0" represented very low quality and "10" indicated very high quality. It included semantic labels (i.e. "low" and "high") at the beginning and end of the scale, and it was presented on a separate screen, to avoid distraction from the image. The observation time was not constrained. Before presenting the stimuli, a white cross was shown at the center of the screen for one second, to force all observations to start from the same fixation at the centre of the screen. This choice was useful to correct possible drifts in calibration.

To limit memory effects, all 54 stimuli were divided in 18 groups of 3 stimuli, sharing both content and distortion type, and so differing only in quality level. The experiment was then structured in 3 sessions, where in each session only one

image out of the 18 groups was selected. The selection was done randomly, but in such a way that each of the 3 images per group was presented in only one of the 3 sessions. We followed this procedure for selecting stimuli per session because we wanted to minimize carry-over effects in the scores between images, which we expected to be bigger for different quality levels than for different contents or distortion types. Between sub-sessions, observers were allowed to rest their eyes for a few minutes. Eventually, the evaluation of the whole dataset, including the breaks between the three sessions, lasted on average 15 minutes.

3. PROCESSING OF EYE TRACKING DATA AND QUALITY SCORES

The experiment produced two main outputs, eye-tracking data and quality scores. To enable their analysis, they were processed as follows.

Eye-tracking equipment records the observer's pupil movements in terms of fixation points and saccades. From these data, saliency maps can be extracted according to various methodologies [14]. As the focus of the experiment was to detect changes in spatial saliency and to retrieve, if existing, their relationship with distortion locations and intensity, no temporal data was included in the analysis. In other words, the fixation points were all equally weighted in the map construction. Saliency maps were obtained for each image I in the dataset by:

1. determining the location of fixation points for each observer separately,
2. averaging all points over all observers to a single fixation map $FM^{(I)}(x,y)$,
3. applying a grey-scale patch to each fixation point in $FM^{(I)}(x,y)$. Patches had a Gaussian intensity distribution with variance, σ, approximating the size of the fovea (about 2° of visual angle). The saliency value, $SS^{(I)}(k,l)$, at location (k,l) of the saliency map for image I (having W_I x H_I pixels) was eventually defined by:

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

where $k \in [1, W_I]$, $l \in [1, H_I]$; (x_j, y_j) are the spatial coordinates of the jth fixation ($j=1 \ldots T$) in the averaged map $FM^{(I)}(x,y)$.

As a result, the saliency map element $SS^{(I)}(k,l)$ reports the probability that location (k,l) of image I gets the average observer's attention. Eventually, 54 maps resulted from the experiment: 9 maps per original image, of which 3 maps corresponded to a blurred version of the original image, 3 to a JPEG compressed version, and 3 to a version distorted with Gaussian noise. In the rest of the paper, we refer to these maps as the Deviated Scene Saliency (DSS) maps.

In addition, for each of the 54 stimuli a Mean Opinion Score (MOS) was computed, averaging the 14 observers' evaluations. When processing these MOS according to the recommendations of the Video Quality Experts Group [15] no outlier observer was found. Nonetheless, a relatively large discrepancy was found among the observers' judgments, resulting on average in a standard deviation on the MOS of 1.35 on a 10-point scale. Figure 2 shows the MOS and its 95% confidence interval for all stimuli.

4. IMPACT OF DISTORTION TYPE AND LEVEL ON IMAGE SALIENCY

Deviations in saliency are analyzed along two factors, namely 1) the dependency on the distortion type and 2) the dependency on the quality level. To analyze deviations in DSS related to the kind of distortion, the stimuli and the corresponding saliency maps are divided into three groups, each consisting of 18 images:

- B: all "blur" images in table I (average quality level 5.56)
- J: all "JPEG" images in table I (average quality level 5.36)
- N: all "noise" images in table I (average quality level 4.68)

To analyze deviations in DSS related to the quality level, the stimuli and the corresponding saliency maps are similarly divided into three groups, each consisting of 18 images:

- LQ: all "low quality" images in table I (average quality level 3.06)

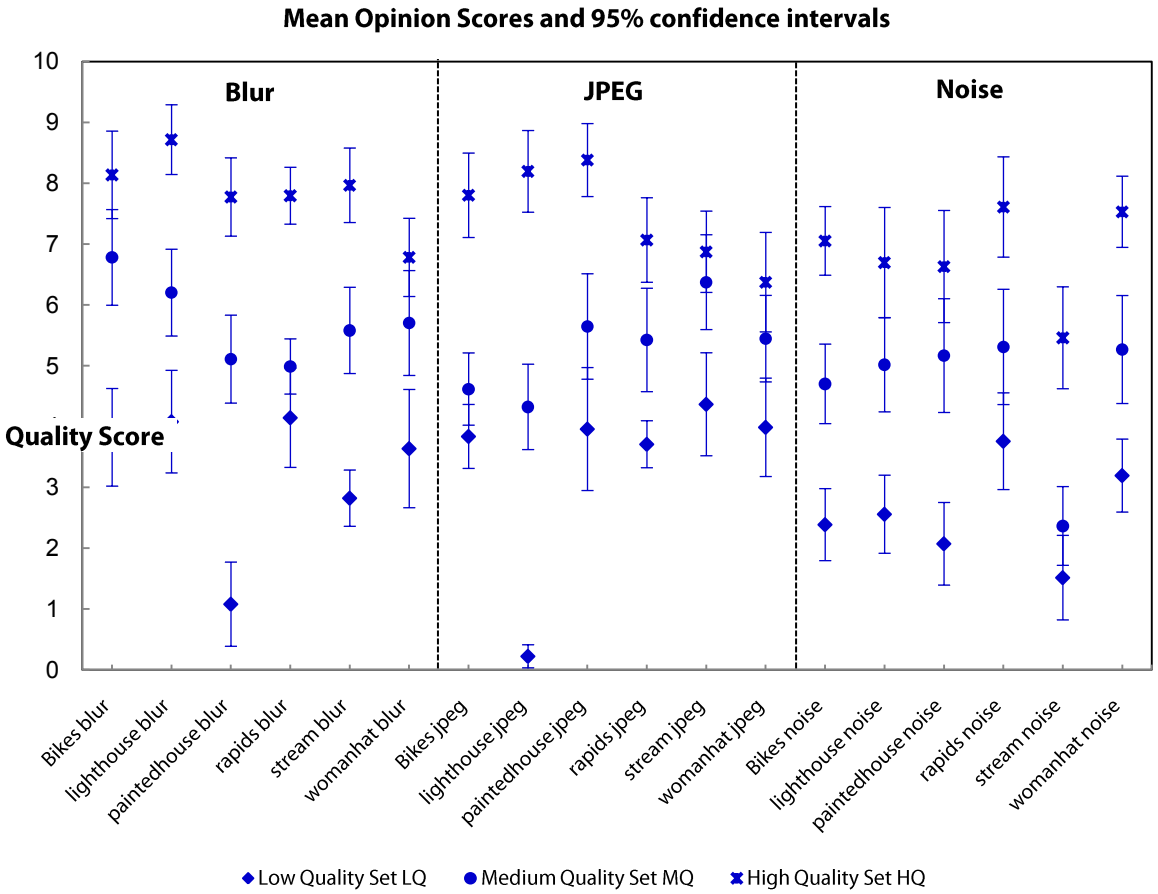


Figure 2 - Mean Opinion Scores for the stimuli in the dataset (cfr. Table I). Each panel in the graph corresponds to a given distortion type: i.e. Gaussian Blur (left), JPEG compression (center) and White Noise (Right). The different symbols in the graph correspond to different quality levels. Error bars indicate the 95% confidence interval of the MOS.

- MQ : all "medium quality" images in table I (average quality level 5.22)
- HQ: all "high quality" images in table I (average quality level 7.38)

The image groups can be visually inspected in figure 2.

Deviations in saliency due to the scoring task are analyzed in two steps. First, the saliency maps obtained during scoring distorted images (DSS) are mutually compared to determine effects of both the distortion type and the quality level. One measure of consistency among saliency maps is the linear correlation coefficient [12], ρ, which quantifies the (symmetrical) strength of the linear relationship between two distributions. The value of ρ ranges between [-1, 1]; 1 indicates maximum linear correlation, while 0 indicates uncorrelated samples. Negative values of ρ indicate inversely correlated samples (the closer to 0, the less correlated). In practice, a higher value of ρ indicates a larger similarity between two saliency maps.

In a second phase, the DSS maps are compared to the NSS maps (obtained in [5]). In this case, we don't look for mutual differences, but rather for deviations of DSS with respect to a reference saliency distribution, i.e., the natural scene saliency. Two more informative measures can be used in this scenario. The Kullback-Leibler divergence (KLD) gives a measure for the dissimilarity between a test distribution and a reference one. As such, it is a positive quantity; it increases with the dissimilarity in the distributions, and KLD = 0 only in case of identical distributions. In this paper, we adopt the NSS as the reference saliency distribution, and we measure the deviation of the DSS from it by computing:

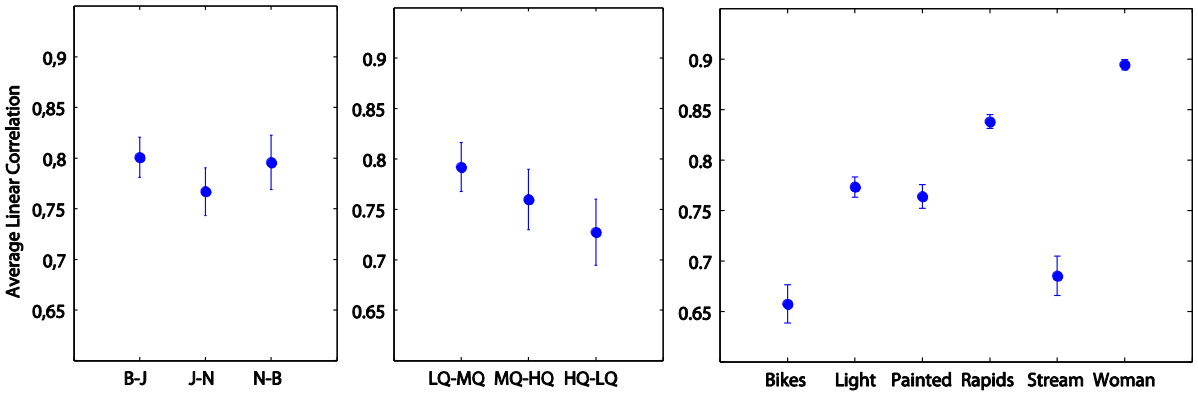
$$KLD\left(DSS^{I}_{d,q}, NSS^{I}\right) = \sum_{x} DSS^{I}_{d,q}(x) \log\left(\frac{DSS^{I}_{d,q}(x)}{NSS(x)}\right)$$
$$x \in I; \quad d \in \{Blur, Jpeg, Noise\}; \quad q \in \{LQ, MQ, HQ\} \tag{2}$$

4.1 Analysis of deviated saliency maps

First, we evaluate differences in DSS resulting from the various distortion types. For the same content and quality level, the correlation in saliency between pairs of one blurred stimulus and one JPEG compressed stimulus (B-J) is computed. This ρ value gives a measure of how, for a given image content and quality level, the saliency distribution changes due to the type of distortion applied to the image. In analogy, the ρ value between the DSS of the JPEG compressed stimulus and the noisy stimulus (J-N), and the ρ value between the DSS of the noisy and blurred stimuli (N-B) are computed. These ρ values are then averaged per group to obtain the values reported in figure 3.a. A similar analysis is repeated for the differences in DSS resulting from different quality levels: for the same content and distortion type, ρ is computed between the DSS of stimuli with low and middle quality (LQ–MQ), with middle and high quality (MQ-HQ), and with low and high quality (LQ–HQ). Figure 3.b shows the corresponding correlation values averaged across image content and distortion types.

Figure 3.a shows how DSS maps obtained for the same quality level, but with different distortions are mutually correlated. The similarity between the various maps is relatively high (ρ∈[0.77, 0.80]), and no distortion seems to produce maps consistently different from those obtained for other distortions. This indicates that all three distortions bring similar saliency distributions as a consequence of scoring the image quality. A more pronounced difference in deviated saliency is found as a function of the quality level of the stimuli (figure 3.b). Maps obtained at LQ and MQ seem to be more correlated than those obtained at LQ and HQ. The main outcome is a significant decrease in correlation between saliency maps with increasing quality levels.

Figure 3.c shows the average of the correlation between all possible pairs of maps for a given content in the dataset (given n = 9 maps per content, we have $n_{pairs} = \frac{(n-1)^2}{2} = 36$ pairs). The graph clearly demonstrates that the correlation between saliency maps is highly dependent on image content. In other words, depending on the specific content, saliency maps corresponding to different quality levels or distortion types may be more or less correlated; e.g., the maps obtained for the various stimuli of the image *Bikes* deviate more among each other than, e.g., the maps obtained for the various stimuli of the image *Woman Hat*, independent on the distortion type or quality level.



(a) *(b)* *(c)*

Figure 3 – Correlation (ρ) among Deviated Saliency maps: comparing various types of distortions averaged over content and quality level *(a)*, comparing various quality levels averaged over distortion type and content *(b)* and comparing various contents averaged over all combinations of distortion type and quality level *(c)*. The error bars represent the standard error of the mean of the correlation values obtained over all pairs.

4.2 Discrepancies between Natural and Deviated Scene Saliency

For the analysis between NSS and DSS we use the KLD value. The average KLD between the NSS map and the DSS map for the B, J and N groups of stimuli (averaged over content and quality level) is given in figure 4.a. Similarly, figure 4.b shows the average KLD value obtained for the LQ, MQ and HQ stimuli (averaged over content and distortion type). Finally, figure 4.c reports the KLD values per image content (averaged over distortion type and quality level), while figure 4.d gives the KLD values per distortion type and quality level averaged over all contents. In addition, the correlation coefficient between the NSS and DSS map is computed for each distortion type (averaged over all contents and quality levels) and for each quality level (averaged over all contents and distortion types). These ρ values are reported in table II and allow us to compare the differences between NSS and DSS with the within-DSS differences (see fig. 3).

Table II – Correlation (ρ) between Natural Scene Saliency and Deviated Saliency maps: depending on the distortion (columns 1 through 3) and on the quality level (columns 4 through 6)

ρ	*Blur*	*JPEG*	*Noise*	*LQ*	*MQ*	*HQ*
NSS	0,71	0,70	0,70	0,69	0,71	0,71

Table II shows that the deviation of DSS from NSS is bigger than the one found between different DSS maps (cfr. fig.3). The KLD analysis gives better insight in how these deviations relate to distortion type and quality level. Results in figure 4 suggest that the deviation in saliency from NSS is more related to the quality level of the image than to the type of distortion affecting it. In particular, the lower the quality, the larger is the deviation. This trend is confirmed by the graph in figure 4.d. The KL divergence between NSS and DSS decreases with increasing quality for each of the three distortion

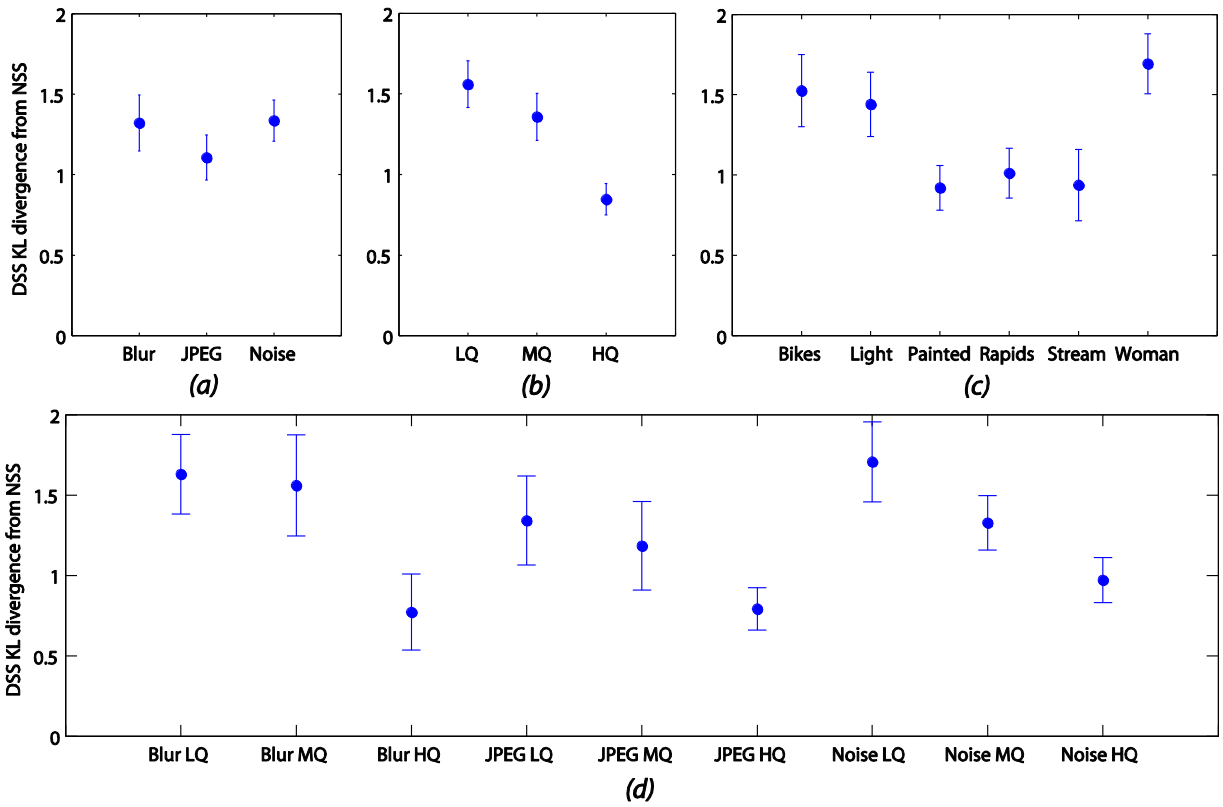


Figure 4 – Kullback-Leibler divergence (KLD) among Deviated Saliency maps: comparing various types of distortions averaged over content and quality level *(a)*, comparing various quality levels averaged over distortion type and content *(b)* and comparing various contents averaged over all combinations of distortion type and quality level *(c)*. Figure *(d)* shows the combined effect of quality level and distortion kind. The error bars represent the standard error of the mean of the correlation values obtained over all pairs.

types separately. The behavior is most evident for blurred and noisy stimuli, but the trend is visible for JPEG distorted stimuli too. Figure 4.c shows that the effect of image content on the change in DSS from NSS is different from what was found when mutually comparing distorted saliency maps (cfr. fig. 3.c).

4.3 Investigating the spatial nature of saliency deviation

The results reported in the previous sections point out that there is a deviation in saliency from NSS as a consequence of asking people to score quality of degraded images, and suggest that these deviations are more influenced by the quality level of the stimulus rather than by the kind of distortion applied to it. So far, however, we did not discuss how the saliency of the image actually spatially changes as a consequence of the scoring task. We now look into how saliency spatially changes both from a global and a local perspective.

Figure 5 presents the saliency maps obtained for free looking (top row) and scoring of different versions of the image content *rapids*. In particular, maps corresponding to two images varying in blur (HQ and LQ) and two varying in compression rate (HQ and LQ) are compared in the middle column. At a first sight, we notice a global reduction of the

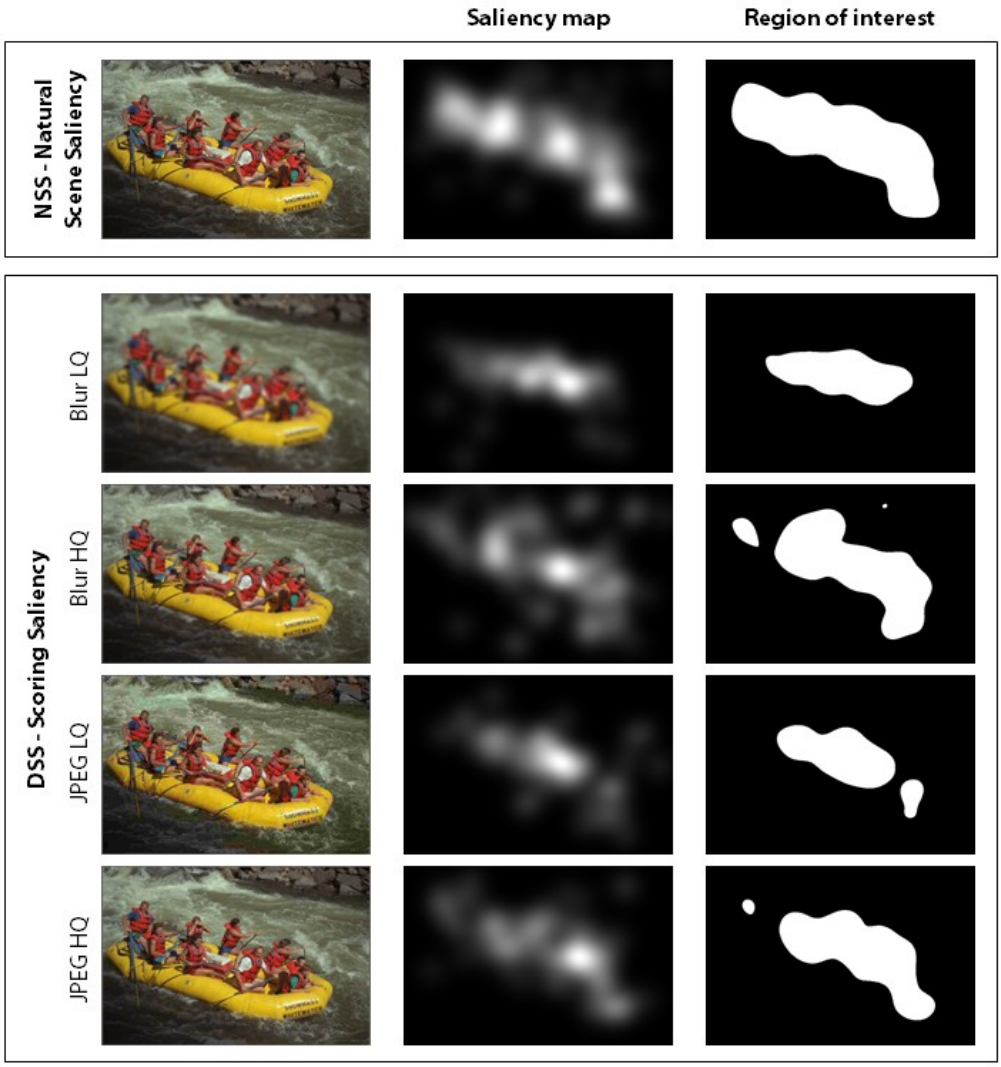


Figure 5 – Comparison between NSS (top row) and DSS, obtained for a High Quality and Low Quality version of a blurred and compressed image of the original content *Rapids*. The middle column shows the saliency maps, whereas the right column illustrates the corresponding ROI, obtained by binarizing the saliency maps with a threshold $SS(k,l) > 0.25$ ($Y > 64$).

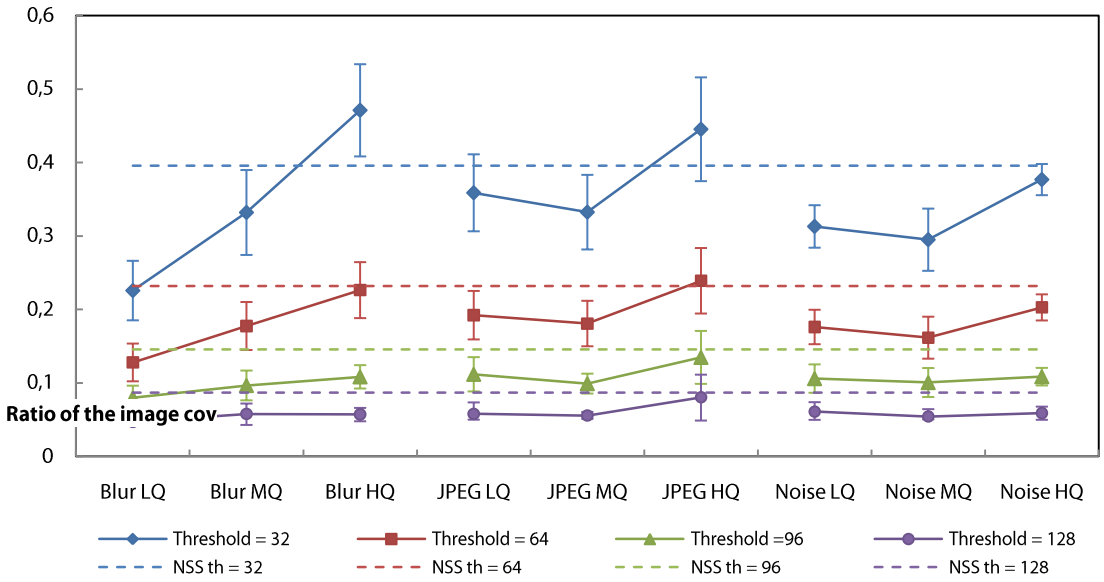


Figure 6 - Combined effect of distortion and quality level on the size of the ROI in the DSS maps. The ratio of the image surface covered by the ROI (averaged over content) is reported for different values of the threshold (expressed in the Y-value of the intensity in the saliency map). The dashed lines represent the average size of the ROI in the NSS map.

spread of saliency over the image as a consequence of scoring low quality images. This tendency is further confirmed when analyzing the ROI, as shown in the right column of Figure 5. This ROI is defined by binarizing the saliency maps with a given threshold value $SS(k,l)$. The binarized image or ROI in Figure 5 (obtained with a threshold value of $SS(k,l) = 0.25$) is clearly smaller for a low quality than for a high quality stimulus.

To quantify this tendency further, the size of the ROI is measured in terms of number of pixels covered, and normalized with respect to the total number of pixels in the image. This procedure is repeated for all NSS and DSS using different threshold values. Figure 6 shows the combined effect of distortion and quality level on the size of ROI. For a threshold value of $SS(k,l) = 0.375$ (i.e., a probability higher than 0.375 that pixel (k,l) is attended, corresponding to a Y-value of 96 in the intensity map) the NSS ROI occupies about 15% of the whole image area. The size of the ROI in DSS tends to shrink with an increasing amount of distortion.

There may be two main reasons for the shrinkage in ROI at decreasing quality level of the image:

(1) the time of observation, and so the number of fixations, may be smaller at lower quality,

(2) the attention may be more spread towards background regions in the image, while not contributing to the ROI because the individually spread fixations do not exceed the threshold.

We indeed observed that, at lower qualities, the number of fixations necessary for the observer to produce a judgment is lower than at high qualities, and so is the time of observation. These quantities are quite well correlated with the MOS of the images, as is shown in table III.

Table III – Pearson's correlation coefficient between the measured Mean Opinion Score, the average image observation time and the average number of fixations during the observation.

Pearson's correlation coefficient	*Observation duration*	*Number of fixations*
Mean Opinion Score	0.69	0.63

A possible explanation is that at lower qualities, observers are quickly confident in judging an image, while, with the increase of the quality level, a more accurate observation is needed to formulate a sound judgment. Hence, the number of fixation points increases. As higher numbers of fixations correspond to an increase of the saliency of the whole image, we can state that, under a quality scoring task with an unconstrained observation time, low quality images are less salient than high quality images.

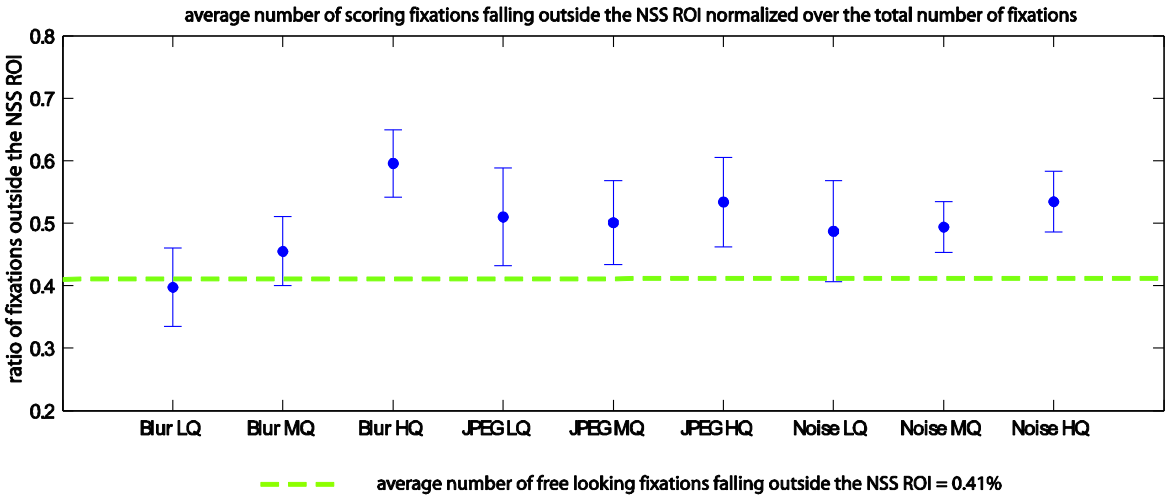


Figure 7 – Average ratio of fixations falling outside the original region of interest (NSS ROI) for images affected by different levels and kinds of distortions, observed under a scoring task. The ROI is identified by setting a threshold of $SS(k,l) = 0.375$. The dashed line represents the average ratio of fixations falling outside the NSS ROI during free looking.

To measure the shift of attention from the ROI to the background, we compute the number of fixation points that, for every DSS, fall outside of the ROI of the corresponding NSS. Figure 7 summarizes the values obtained, normalized over the total number of fixations, for an ROI threshold of $SS(k,l) = 0.375$. The effect of quality level is not so very obvious, except for the blurred stimuli. On average, a large part of the fixation points (50.11%, averaged over all stimuli) falls outside the ROI. With respect to the NSS, this represents an increase of about 20% of the number of fixations outside the ROI. Attention during scoring is thus more spread across the image, as typically observers scan image details before having enough confidence to express a judgment. Eventually, observations are still highly concentrated in the NSS ROI, but a shift of saliency happens from foreground to background. This may be a complementary factor in the observed ROI shrinking.

5. CONCLUSIONS

Saliency maps obtained during quality assessment of distorted images differ among each other, and these differences seem to be more related to the quality level of the related stimulus than to the kind of distortion affecting that stimulus. More prominent changes in saliency are found when comparing the various DSS maps to the NSS map. In this case, again the deviation seems to be most related to the amount of distortion present in the stimulus. In particular, the lower the quality, the more dissimilar is the DSS from the NSS. This structural change may originate from the fact that observers check images more carefully with an increase of the quality level, which is also confirmed by the fact that observation times, number of fixations per image and MOSs are well-correlated. The saliency deviation manifests itself through a shrinking of the ROI and a shift of attention towards the background of the image. While the shrinking phenomenon is again proportional to the amount of distortion for all the distortions considered, the shift of attention seems to depend on the quality level only for blurred images.

In conclusion, the experiment proved that the quality assessment task modifies the natural saliency of images. Therefore, when modeling quality, the masking effect of saliency for distortions in background regions may be less prominent than in a free looking situation. When modeling interactions between visual attention and image quality perception for objective quality assessment, the effect of the quality level of the image on the saliency should be taken into account to design a more accurate integration strategy.

REFERENCES

[1] Yarbus, A., “Eye movements and vision,” New York: Plenum Press (1967).
[2] Antes, J.R., “The time course of picture viewing,” Journal of Experimental Psychology, 103, 62-70 (1974).

[3] Desimone, R., and Duncan, J., "Neural mechanisms of selective visual attention," Annual Review of Neuroscience, 18, 193–222 (1995).
[4] Wang, Z., and Bovik, A. C., "Modern Image Quality Assessment," Synthesis Lectures on Image Video & Multimedia Processing, Morgan & Claypool Publishers (2006).
[5] Liu, H., and Heynderickx, I., "The added value of visual attention in objective quality metrics," Proc. IEEE ICIP 2009, 3097-3100 (2009).
[6] Ninassi, A., Le Meur, O., Le Callet, P., and Barba, D., "Does where you gaze your attention on an image affect your perception of quality? Applying Visual Attention to image quality metrics," Proc. IEEE ICIP 2007, 2, 169-172 (2007).
[7] Larson, E. C., Vu, T. and Chandler, D. M., "Can visual fixation patterns improve image fidelity assessment?," Proc. IEEE ICIP 2008, 3, 2572-2575 (2008).
[8] Redi, J., Liu, H., Gastaldo, P., Zunino, R., and Heynderickx, I., "How to apply spatial saliency into objective metrics for JPEG compressed images?," Proc. IEEE ICIP 2009, 961-964 (2009).
[9] Sadaka, N. G., Karam, L. J., Ferzli, R., and Abousleman, G. P., "A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling," Proc. IEEE ICIP 2008, 369–372 (2008).
[10] Sheikh, R., Wang, Z., Cormack, L., and Bovik, A., "LIVE Image Quality Assessment Database," http://live.ece.utexas.edu/research/quality
[11] 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).
[12] Vu, T., Larson, E. C., and Chandler, D. M., "Visual Fixation Patterns when Judging Image Quality: Effects of Distortion Type, Amount, and Subject Experience," Proc. IEEE SSIAI 2008, 73-76 (2008).
[13] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," Geneva (2002).
[14] De Ridder, H., "Cognitive Issues in image quality measurement," J Electronic Imaging, 10(1), 47-55, (2001).
[15] Keelan, B., and Urabe, H., "ISO 20462, A psychophysical image quality measurement standard," Proc. SPIE 5294, 181-189, (2004).
[16] Wooding, S., "Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps," Behavior Research Methods Instruments and Computers, 34(4), 509-517 (2002).
[17] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," http://www.vqeg.org/, (2000).
[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE* Transactions on Image Processing, 13(4), 600-612 (2004).