Copyright © IEEE.
Citation for the published paper:

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of BTH's products or services Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by sending a blank email message to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

VISUAL FIXATION PATTERNS IN SUBJECTIVE QUALITY ASSESSMENT: THE RELATIVE IMPACT OF IMAGE CONTENT AND STRUCTURAL DISTORTIONS

Ulrich Engelke[†], Hans-Jürgen Zepernick[†], and Anthony Maeder^{*}

[†]Blekinge Institute of Technology, 371 79 Karlskrona, Sweden, E-mail: uen@bth.se

^{*}University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia

ABSTRACT

The viewing behaviour of human observers during image quality assessment is analysed. In this respect, the relative impact of image content and structural distortions is of particular interest. Two subjective experiments were conducted as a basis for analysis; a region-of-interest (ROI) experiment and an eye tracking experiment under quality assessment task. A correlation analysis and a receiver operating characteristics analysis reveal, that quality assessment takes place mainly within the ROI, indicating a higher impact of the content as compared to the distortions. This seems to be only contradicted by very strong and localised distortions outside the ROI.

Index Terms— Eye tracking, region-of-interest, image quality assessment, visual attention.

1. INTRODUCTION

Subjective quality experiments serve the design and validation of image quality metrics [1]. The quality scores are typically analysed using statistical measures and are averaged into mean opinion scores (MOS), which are used as a ground truth for quality metric design. As useful as these scores are, they provide very limited information with respect to the assessment strategies that human observers deploy to reach a quality judgement. To gain more insight, it is of great interest to collect more information during the quality assessment.

One non-intrusive way to collect such information is by means of eye-tracking which is considered to reflect well the visual attention (VA) of human observers [2]. We therefore analyse the viewing behaviour of human observers during image quality assessment by recording the gaze patterns using an eye tracker. In this context, we specifically focus on the relative impact of structural distortions and of the perceived level of interest in natural image content. For this purpose we conducted two subjective experiments. The first one focused on the identification of regions-of-interest (ROI) in a number of images. These reference images were used to create a large set of distorted images using a wireless link simulation model. In the second experiment, the distorted images were shown to different observers who rated the quality while we recorded their gaze patterns. The results of the two experiments enable

Fig. 1. Test images showing different artifacts: (a) blocking, (b) ringing, (c) ringing and block intensity shifts.

us to analyse the impact of the level of interest in an image and the structural distortions on the viewing behaviour.

The paper is organised as follows. Section 2 explains the two experiments we conducted. Section 3 provides a detailed analysis of the results. Conclusions are drawn in Section 4.

2. SUBJECTIVE EXPERIMENTS

The subjective ROI experiment we conducted is referred to as SE_{ROI} and the eye tracking experiment under quality assessment task is referred to as SE_{ET} . Both experiments and the creation of the test images are introduced in the following.

2.1. Test images

We used seven well known images in both experiments SE_{ROI} and SE_{ET} , namely, 'Barbara' (B), 'Elaine' (E), 'Goldhill' (G), 'Lena' (L), 'Mandrill' (M), 'Peppers' (P), 'Tiffany' (T). Based on these reference images we created a set of 80 distorted images using a simulation model of a wireless link. To be precise, the images were compressed in JPEG format followed by a (31,21) BCH channel code for error protection. Binary phase shift keying was used as modulation technique and a Rayleigh flat fading channel with additive white Gaussian noise was implemented to simulate the wireless channel.

The resulting test images contained a range of structural distortions, including blocking, blur, ringing, intensity shifts, and combinations thereof. The test images are discussed in detail in [3] and some example images are shown in Fig. 1.

Fig. 2. Mean ROI for the images (left to right): 'Barbara', 'Elaine', 'Goldhill', 'Lena', 'Mandrill', 'Peppers', 'Tiffany'.

2.2. Region-of-interest selection

The aim of subjective experiment SE_{ROI} was to identify ROI in the seven reference images. For this purpose, thirty observers were asked to select a single region that was of particular interest to them in each of the images. No constraints were imposed regarding the size of the ROI, however, for simplicity only rectangular shaped ROI were considered. The thirty ROI for each image were then transferred into mean ROI, which are shown in Fig. 2. Here, the mean location and mean dimension of all ROI selections were determined by averaging, respectively, the ROI center coordinates and the ROI dimensions in both horizontal and vertical direction. The ROI experiment is explained and analysed in detail in [4].

2.3. Eye tracking under quality assessment task

The procedures of experiment SE_{ET} were designed according to ITU-R Rec. BT.500-11 [5]. The seven reference and eighty distorted images were shown to fifteen observers who were instructed to rate the quality on a 5-point scale. The images were presented about 8 seconds each in pseudo-random order in two consecutive sessions. The reference images have been presented in both sessions as hidden references.

An EyeTech TM3 [6] eye tracker was used throughout the experiment to record the gaze patterns of the observers at a frequency of 50 Hz. The recorded gaze patterns were then post-processed into visual fixation patterns (VFP) and saliency maps (SM) [7] for further analysis. To minimise the impact of the quality rating and the gaze patterns on each other, the observers were asked to do the quality scoring during a 5 second mid-grey screen between two consecutive images.

3. ANALYSIS

In the following, we analyse the impact of the distortions and the image content on the viewing behaviour. The former is addressed by means of correlation analysis. The latter is evaluated using receiver operating characteristics (ROC) analysis.

3.1. Viewing consistency between the hidden references

The impact of the distortions on the viewing behaviour is quantified with the Pearson linear correlation coefficient, ρ_P ,

Table 1. Correlation coefficient, ρ_P , between the SM of the reference images from the first and second session.

	Barbara	Elaine	Goldhill	Lena	Mandrill	Peppers	Tiffany
ρ_P	0.973	0.978	0.946	0.952	0.914	0.912	0.966

between the SM of the distorted images and the corresponding reference images. In this respect, a higher correlation corresponds to a higher similarity between the SM. As each reference image has been presented twice during the experiment, once in the first session and once in the second session, we can first evaluate the consistency of the observers' viewing behaviour when being presented the same image twice.

The correlation coefficients ρ_P between the SM from the first and second session are given in Table 1. One can see that for all reference images, the correlations are well above 0.9. In fact, the images 'Barbara', 'Elaine', 'Lena', and 'Tiffany' even exhibit correlations above 0.95 between the SM. These images contain humans and their faces which are of high interest to the observers, as can be seen from Fig. 2. Given the high correlations in Table 1, the viewing behaviour can be considered to be fairly consistent on the reference images. As such, changes in the SM, and thus lower correlations, can be related to changes in the image content in terms of distortions.

3.2. Visual attention to structural distortions

The correlations of all distorted images with their respective reference images are presented in Fig. 3 (with the capital letters in the legend denoting the respective reference images). In addition, the average and standard deviation over all correlations related to a particular image content are summarised in Table 2. From Fig. 3 one can see that there is a wide range of correlations, indicating, that the different distortions changed the viewing behaviour to different degrees. The mean correlations, μ_{ρ} , in Table 2 further reveal that this change depends on the image content, as the correlations are considerably lower for 'Barbara', 'Mandrill', and 'Peppers'.

We attempted to determine quantitative relationships between the distortions contained in an image and the related SM computed from the gaze patterns of all observers. How-

Fig. 3. Correlation coefficient ρ_P between the SM of all distorted images and their corresponding reference images.

Table 2. Mean and standard deviation of ρ_P between the SM over all distorted images and reference images.

	Barbara	Elaine	Goldhill	Lena	Mandrill	Peppers	Tiffany
μ_ρ	0.87	0.969	0.904	0.931	0.81	0.868	0.923
σ_ρ	0.045	0.022	0.055	0.038	0.069	0.051	0.035

ever, the relationship between these two factors seems highly complex and not as intuitive as one might expect. One would, for instance, expect highly distorted images to disturb the gaze patterns more as compared to slightly distorted images. This, however, has been found not to be true. For this reason, we briefly summarise some qualitative observations:

1. **Distortion location:** The observers generally tend to analyse the quality within the average ROI obtained from experiment SE_{ROI} . This is particularly true when distortions are present both inside and outside the ROI. If distortions are only present outside the ROI, then the search range is shifted to other parts of the image.
2. **Distortion distribution:** Global distortions generally do not alter the SM as much as local distortions do. Local distortions are more likely to change the gaze patterns, in particular if they are clearly visible and located outside the ROI.
3. **Distortion strength:** The distortion strength has comparably lower impact on the alteration of the SM, as compared to the distortion location and distribution. Strong distortions that are locally distributed tend to change the SM but so do weak locally distributed distortions. In fact, it seems that subtle distortions often change the SM more than strong distortions which can be attributed to the fact that they need to be attended relatively longer for thorough analysis.

4. **Distortion type:** The different distortion types in relation to their strengths seem to have a considerable impact on the viewing behaviour. Blocking artifacts usually need to be fairly strong for the observers to attend them. On the other hand, ringing artifacts can considerably change the SM even if they are only very subtle which may be attributed to the more complex structure of the ringing artifacts that leads to a more thorough analysis. Block intensity shifts are typically analysed at the border between the two different intensities rather than in either of the intensity shifted area.

3.3. Visual attention in the region-of-interest

The degree to which the quality assessment takes place within and outside the ROI is estimated using ROC analysis [8] between the ROI from experiment SE_{ROI} and the SM from experiment SE_{ET} . ROC analysis is typically used for binary classification of a performance measure into one of two classes, a positive class and a negative class. Here, we define SM pixels inside the ROI to belong to the positive class and SM pixels outside the ROI to belong to the negative class. Given our earlier observations that quality assessment seems to take place in the ROI, we would expect the high magnitudes in the SM to be within the ROI.

The outcomes of the ROC analysis are twofold; firstly, the ROC curve visualises the relative amount of saliency that has correctly been classified to belong to the ROI, in terms of the true positive rate (TPR), over the saliency points that were wrongly classified to belong to the ROI, the false positive rate (FPR). Curves in the ROC space that are located towards the upper left corner indicate good separability between the classes, whereas ROC curves towards the diagonal of the ROC space represent poor separability. The area under the ROC curve (AUC) then quantifies the classification performance in a range from 0 to 1, with a higher value indicating that more saliency has been present in the ROI.

The ROC curves for all distorted images are presented in the Fig. 4 in separate plots for each of the seven contents. The related mean AUC μ_{AUC} over all distorted images of each content are presented in Table 3, along with the corresponding standard deviations σ_{AUC} . It can be seen that for all distorted images, the ROC is located far in the upper left corner which reveals that the quality assessment indeed mainly takes place in the ROI. This is further supported by the high AUC values for all contents. The somewhat lower AUC for the 'Barbara' image can be attributed to the ROI being in the periphery of the image, however, the AUC are still very high.

3.4. Impact of quality assessment duration

From analysing the gaze patterns, we found that the degree to which the quality is performed in the ROI changes with the duration of the quality assessment. To be more precise,

Fig. 4. ROC curves between the SM and ROI for all distorted images of (left to right): 'Barbara', 'Elaine', 'Goldhill', 'Lena', 'Mandrill', 'Peppers', and 'Tiffany'.

Table 3. Mean, μ_{AUC} , and standard deviation, σ_{AUC} , of the AUC between the ROI and the SM of all distorted images.

	Barbara	Elaine	Goldhill	Lena	Mandrill	Peppers	Tiffany
μ_{AUC}	0.818	0.97	0.945	0.946	0.907	0.884	0.97
σ_{AUC}	0.029	0.01	0.025	0.031	0.038	0.039	0.018

Table 4. AUC for the ROI selections and the SM created from the first, F_1 , and last fixation, F_L , of each observer.

	Barbara	Elaine	Goldhill	Lena	Mandrill	Peppers	Tiffany
F_1	0.943	0.98	0.945	0.974	0.871	0.915	0.962
F_L	0.812	0.959	0.934	0.926	0.852	0.834	0.929

after appearance of the image, the observers tend to consult the ROI for quality assessment. Towards the end of the image presentation time, however, the gaze shifts to some degree to other parts of the image, outside the ROI.

To provide quantitative evidence for this phenomenon, we created SM based on only the first fixation of every observer, F_1 , and also based on only the last fixation of every observer, F_L . The AUC computed between these SM and the ROI are presented in Table 4. It can be seen that for all seven contents, the AUC is larger for the SM based on the first fixations, F_1 , as compared to the SM based on the last fixations, F_L . Thus, the quality assessment seems to generally start in the ROI and then shifts partly to other locations in the image, with the ROI still dominating though. This is illustrated by the example images in Fig. 5 which show the first and last fixations of all observers in the top and bottom row, respectively.

4. CONCLUSIONS

We analysed the relative impact of structural distortions and image content on the quality assessment strategy of human observers. Qualitative analysis revealed that the distortion location, distribution, type, and strength impact on the viewing behaviour. However, the distortion seems to play a minor role compared to the image content, and in particular in relation to the perceived level of interest.

Fig. 5. First (top row) and last (bottom row) fixation of every observer for the images 'Barbara', 'Goldhill', and 'Tiffany'.

5. REFERENCES

- [1] S. Winkler, *Digital Video Quality - Vision Models and Metrics*, John Wiley & Sons, 2005.
- [2] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, pp. 192–203, 2001.
- [3] U. Engelke, T. M. Kusuma, H.-J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 525–547, July 2009.
- [4] U. Engelke and H.-J. Zepernick, "A framework for optimal region-of-interest based quality assessment in wireless imaging," *Journal of Electronic Imaging, Special Section on Image Quality*, vol. 19, no. 1, Jan. 2010.
- [5] International Telecommunication Union, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500-11, ITU-R, 2002.
- [6] EyeTech Digital Systems, "TM3 eye tracker," <http://www.eyetechds.com/>, 2009.
- [7] U. Engelke, A. J. Maeder, and H.-J. Zepernick, "Visual attention modelling for subjective image quality databases," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, Oct. 2009, pp. 1–6.
- [8] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.