ASSESSING THE QUALITY OF COMPRESSED IMAGES USING EEG

Lea Lindemann, Marcus Magnor

Computer Graphics Lab, TU Braunschweig, Germany

ABSTRACT

The way images are perceived by a human observer is becoming increasingly important in visual media, e.g. for (photo-) realistic image synthesis or memory-efficient encoding of large volumes of image/video data. Traditionally, perceived image quality is assessed by either using specially developed image quality metrics or by conducting user studies. In this paper, we investigate the use of electroencephalography as a tool for evaluating image quality. We demonstrate that the presence of artifacts reliably elicits a measurable response in the brain. We furthermore show that the reaction varies depending on the severity of artifacts, so that it may be used in order to objectively quantify image quality.

***Index Terms*—** Electroencephalography, image quality

1. INTRODUCTION

In recent years, human visual perception has become increasingly important in digital image research. Exemplary applications include the creation of (photo-) realistic images, and encoding and compression of image/video data with minimal effect on perceived image quality. Different methods have been applied to assess image quality [1]. On the one hand, different quality measures based on knowledge about human visual perception have been developed [2]. On the other hand, user studies are increasingly used to assess the advantages of newly developed algorithms and compare them to existing techniques [3].

Our work is focusing on the question whether the impact of artifacts on perceived image quality can be assessed using electroencephalography. An electroencephalogram (EEG) is the recording of ongoing brain activity, called *spontaneous EEG*, in which is embedded the activity produced by the brain's processing of stimuli. Is a stimulus repeatedly presented to an observer under controlled conditions, this stimulus-specific brain activity, called *event-related potential* (ERP), can be extracted by means of averaging [4]. One advantage of this approach over conventional user studies could be that it may prove to be an objective measure of perceived image quality degradation.

As a first step to find out if the assessment of image quality with EEG is a feasible method, we examine the brain's response to JPEG-compressed images. JPEG-encoded images can suffer from artifacts which reduce the perceived image quality, depending on the compression ratio. We chose JPEG because it is the predominantly used lossy image encoding standard.

This work has been funded by the European Research Council ERC under contract No. 256941 "Reality CG".

Fig. 1. The test images used: *baboon*, *peppers*, and *flower*.

2. RELATED WORK

A wide variety of image quality metrics exists. Some, like mean-square-error (MSE) or peak signal-to-noise ratio (PSNR), evaluate the quality of an image by comparing it to a reference image using statistical or correlation-based methods [1], which, however, do not always correlate well with perceived image quality [5]. Perceptually motivated methods adapt properties of the human visual system (HVS) [2]. In [1], for example, nonlinear perception of luminance, contrast sensitivity, and masking effects based on spatial frequencies are built into a complex model of the low-level HVS. In [6], a top-down approach implements the supposed functionality of the HVS by locally comparing pixel intensity patterns after normalization for luminance and contrast. Other approaches evaluate the aesthetic quality of images and even videos using features based on photography techniques like composition and depth of field [7].

There has not been much research regarding the evaluation of image quality using EEG measurements. More often, EEG is used for brain computer interface (BCI) applications, where it assists users in image classification tasks [8]. Hayashi et al. used EEG data to evaluate the quality of high resolution images, assuming that images with good quality produce a higher amount of alpha-waves than images with bad quality [9]. We, on the other hand, focus on ERPs that are the brain's reaction to a specific stimulus.

3. EXPERIMENT

3.1. Participants

Ten right-handed individuals (2 female, 8 male), age 21 - 27, with normal or corrected-to-normal vision participated in the experiment. All subjects had average experience with digital footage and no involvement in professional image/video editing or rendering.

3.2. Stimuli

Three 512×512-pixel JPEG-images were used as test material Fig. 1. Each image was presented in 7 different states of compression: The original image (quality setting 100, i.e. no compression) and 6 compressed images with quality setting 84, 68, 52, 36, 20, and 4 were presented.

One trial consisted of 4 stimuli: A fixation screen (f_1), the uncompressed image (i_1), instantly followed by one of the 7 compressed versions of the image (i_2), and then another fixation screen (f_2). The stimuli f_1, i_1, and i_2 each were shown for $500ms$, whereas f_2 was terminated by the response of the participant pressing a button, which also started the next trial. The fixation screen separated individual trials and consisted of a gray background with a white dot in the center. The fixation dot stayed visible while images were displayed in front of the gray background during intervals i_1 and i_2. The participants were seated so that the images covered $5.5° \times 5.5°$ of visual angle, i.e. the whole image was (at least peripherally) visible without eye movement.

3.3. Procedure

The participants were asked to fixate the white dot and report after the image had disappeared, i.e. during f_2, whether they had detected a change during the period i_1, i_2. The response was given by clicking the left or right mouse button. During one session, the participant was shown trials for one of the three test images. Each of the three test images was shown to 5 different participants. The 7 compressed images, used in i_2, were blockwise pseudo-randomized, so that the same compression level was not shown on successive trials. All 7 compression levels were shown once before a repetition occurred. All compressed images occurred equally often.

A second, identical series of tests, using the quality settings 100, 52, 44, 36, 28, and 12 for i_2, provided additional data for a better signal-to-noise-ratio at higher image quality settings and new data for other quality settings. The image with quality setting 100 was shown twice as often as the other compression ratios to keep the change detection rate at approximately the same level as in the previous configuration.

3.4. Data acquisition and processing

During the experiment the ongoing EEG was recorded from 32 scalp sites, according to the international 10-20 system. Additionally, horizontal and vertical electrooculogram

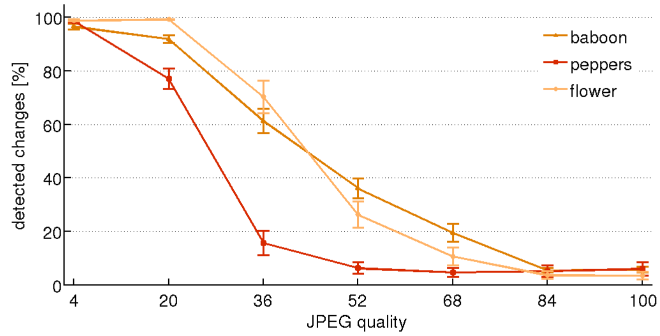


Fig. 2. The percentage of detected changes reported by the participants for different quality settings during the EEG recording (100 = no compression).

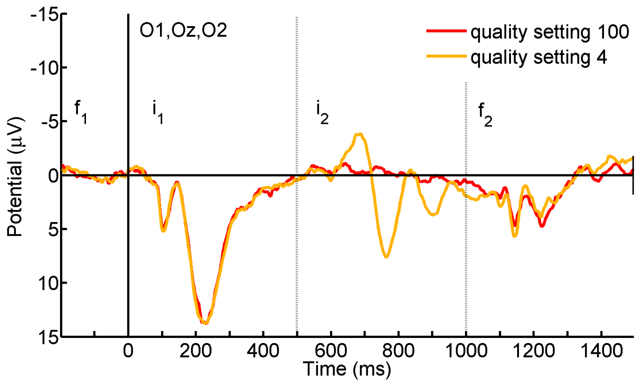
(EOG) were recorded, as well as the EEG at both mastoids as reference. The responses were logged and simultaneously counted.

After recording, the acquired data was high pass-filtered with a cutoff frequency of 0.5 Hz to remove drifts and DC offset. After epoching, trials with artifacts (blinks, saccades,...) were rejected semi-automatically. The remaining trials were sorted by stimulus i_2 and response. For each group, all trials from all participants were averaged to reduce the signal-to-noise ratio and extract ERPs from the spontaneous EEG [4].

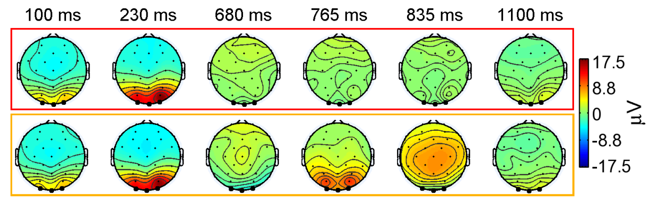
4. RESULTS

Fig. 2 shows the percentage of detected changes for all quality settings of i_2. As expected, the detection rate increases with increasing compression, since more obvious artifacts appear. The psychometric functions for the scenes *baboon* and *flower* are quite similar, whereas the psychometric function for *peppers* suggests that it is harder to detect artifacts in this scene. This may be due to the lesser amount of high frequency content. Error bars show that on intermediate qualities the number of detection varies stronger between subjects than for very high or low quality.

The evaluation of the ERPs reveals that the brain reacts to JPEG artifacts, and that the reaction varies with compression ratio. All subsequent images show ERPs recorded for the scene *flower*, but the responses are very similar for the other test images. Times are specified relative to the beginning of i_1. Fig. 3a illustrates the difference between ERPs elicited by a no-change stimulus, for which the participants did not report visible artifacts, and a maximum-change stimulus, for which the participants did report noticable artifacts. At the beginning of the trial the ERPs are identical, as the same image i_1 is processed. If i_1 and i_2 are identical, no event-related activity occurs after the initial image i_1 has been processed. However, there is considerable activity visible for the maximum-change stimulus when i_2 is at lowest quality setting. After the onset of f_2 the ERPs again align, since the same stimulus is processed. In addition to the ERPs, Fig. 3b shows topographic



(a) Trial-averaged EEG activity



(b) Topographic scalp maps

Fig. 3. The ERPs (a) show the activity, averaged over occipital electrodes O1, Oz, and O2, for the no-change stimulus (red), when i_1 and i_2 are identical (quality setting 100), and the maximum-change stimulus, when i_2 has quality setting 4 (orange). The topographic maps in (b) show the difference in voltage deflections over the scalp at different points in time for the no-change stimulus (red box) and the maximum-change stimulus (orange box).

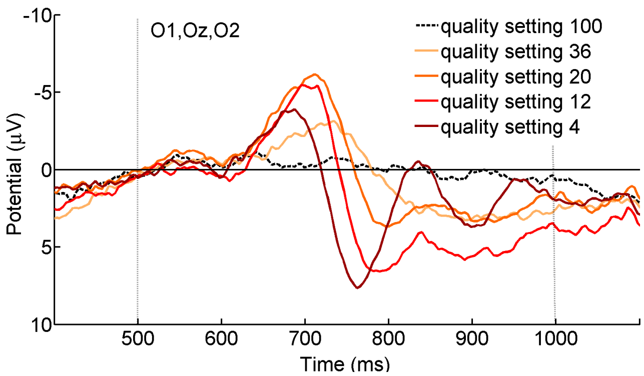


Fig. 4. The ERP averaged over electrodes O1, Oz, O2 for different image quality settings. For higher image quality (less compression) changes are so seldom detected that the resulting ERPs exhibit a very low signal-to-noise ratio.

scalp maps that illustrate the spatial layout of the voltage deflections for different points in time.

Fig. 4 shows the difference in ERPs for varying compression ratios, focusing on interval i_2. For the lowest quality setting, an oscillation with decreasing amplitude phase locked to the stimulus can be observed. With increasing quality the oscillation subsides and becomes a positive deflection that finally converges towards the no-change ERP. A problem with compression ratios lower than the depicted is that the number of trials in which the participants detected a change is very low so that resulting ERPs have a low signal-to-noise ratio, if they can be extracted at all. Still, the collected data show the trend described above. This indicates that it is possible to compare image quality based on ERPs, and to determine the approximate quality perception of an image by means of the ERP.

Another issue with ERPs in general is, that they can vary depending on experimental parameters. For example, the frequency with which a certain stimulus is detected may influence the amplitude of an ERP considerably. Therefore, it is preferable to distinguish ERPs by their shape rather than by absolute values like amplitude. In our experiment, brain activity averaged over the electrode sites O1, Oz, and O2 worked best. Fig. 5 a-d show the averaged activity at other combined electrode sites along the head's midline. While occipital electrodes exhibit different ERP shapes, responses to all compression ratios look very similar at other sites, apart from the ERP elicited at frontal and central electrodes by the maximum-change stimulus. Figures 3 and 4 show ERPs averaged over occipital sites O1, Oz, and O2.

Fig. 6 illustrates that the reaction to the highest compression ratio is very similar for all test images, even if the reaction to the uncompressed image i_1 differs. This shows that the brain's response to artifacts is independent of scene content. With decreasing compression, ERP similarity decreases slightly, but this may also be caused by the difference in signal-to-noise ratio. For the scene *peppers* the number of noticable changes decreases faster than for the other two scenes (Fig. 2), which leads to a worse signal-to-noise ratio.

5. CONCLUSIONS AND FUTURE WORK

We showed that the presence of JPEG artifacts elicits a response in the brain, measurable with EEG. The effect varies with the quality setting and can best be observed at occipital sites. This indicates that it may be possible to examine image quality using EEG. In future work, we want to find out how precisely and reliably differences in image quality can be determined by ERPs, and if the shape of ERPs can be estimated based on already known ERPs. A starting point will be to approximate the known ERPs by a surface, as shown in Fig. 7, and examine if newly measured ERPs match the surface at the

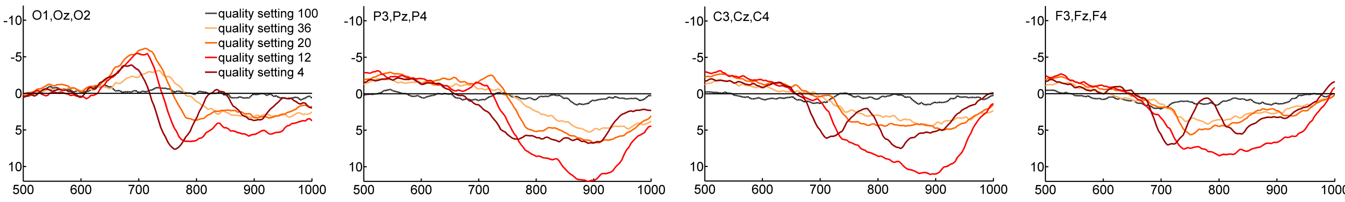
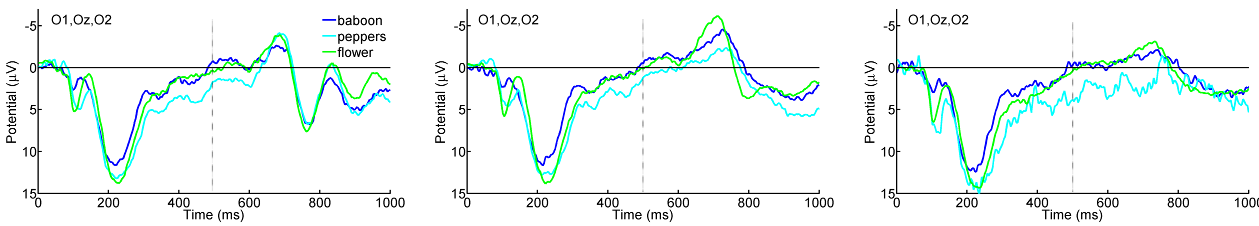


Fig. 5. ERPs for the same compression ratios as in Fig. 4, averaged at occipital (O1, Oz, O2), parietal (P1, Pz, P2), central (C1, Cz, C2), and frontal (F1, Fz, F2) electrode sites.



(a) quality setting 4 (b) quality setting 20 (c) quality setting 36

Fig. 6. The reaction to low i_2 image quality is very similar for all scenes. For better quality the similarity decreases. Whether this is caused by a smaller signal-to-noise ratio or a weaker reaction requires further investigation.

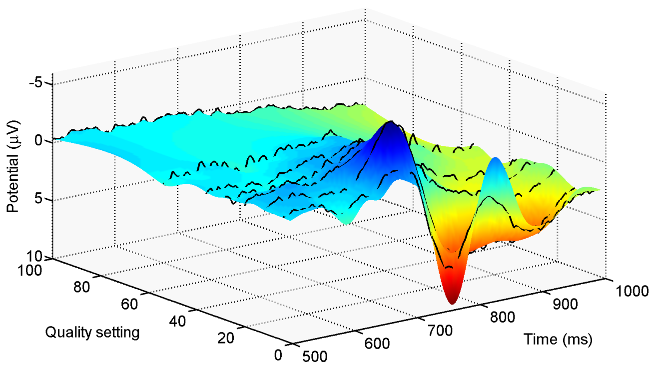


Fig. 7. An exemplary surface approximation for the measured ERPs (black lines).

according compression. Another issue that needs more investigation is the similarity between ERP responses to different images.

6. REFERENCES

[1] A.M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," *Communications, IEEE Transactions on*, vol. 43, no. 12, pp. 2959 –2965, Dec. 1995.

[2] U. Engelke and H.-J. Zepernick, "Perceptual-based quality metrics for image and video services: A survey," in *Next Generation Internet Networks, 3rd EuroNGI Conference on*, May 2007, pp. 190 –197.

[3] K. Berger, C. Lipski, C. Linz, A. Sellent, and M. Magnor, "A ghosting artifact detector for interpolated image quality assessment," in *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, New York, NY, USA, 2009, APGV '09, pp. 128–128, ACM.

[4] S.J. Luck, *An introduction to the event-related potential technique*, MIT press, Cambridge, MA, 2005.

[5] S. Daly, "Digital images and human vision," chapter The visible differences predictor: an algorithm for the assessment of image fidelity, pp. 179–206. MIT Press, Cambridge, MA, USA, 1993.

[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 –612, 2004.

[7] Yiwen Luo and Xiaoou Tang, "Photo and video quality evaluation: Focusing on the subject," in *Proceedings of the 10th European Conference on Computer Vision: Part III*, Berlin, Heidelberg, 2008, ECCV '08, pp. 386–399, Springer-Verlag.

[8] A.D. Gerson, L.C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 174 –179, 2006.

[9] H. Hayashi, H. Shirai, M. Kameda, S. Kunifuji, and M. Miyahara, "Assessment of extra high quality images using both eeg and assessment words on high order sensations," in *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, 2000, vol. 2, pp. 1289 –1294 vol.2.