Studying the Effect of Optimizing the Image Quality in Saliency Regions at the Expense of Background Content

Hani Alers[a], Hantao Liu[a], Judith Redi[b], Ingrid Heynderickx[ac]
[a]Delft University of Technology, Delft, The Netherlands;
[b]The University of Genoa, Genoa, Italy;
[c]Philips Research Laboratories , Eindhoven, The Netherlands

ABSTRACT

Manufacturers of commercial display devices continuously try to improve the perceived image quality of their products. By applying some post processing techniques on the incoming image signal, they aim to enhance the quality level perceived by the viewer. Applying such techniques may cause side effects on different portions of the processed image. In order to apply these techniques effectively to improve the overall quality, it is vital to understand how important quality is for different parts of the image. To study this effect, a three-phase experiment was conducted where observers were asked to score images which had different levels of quality in their saliency regions than in the background areas. The results show that the saliency area has a greater effect on the overall quality of the image than the background. This effect increases with the increasing quality difference between the two regions. It is, therefore, important to take this effect into consideration when trying to enhance the appearance of specific image regions.

Keywords: Image quality, saliency, region of interest, eye tracking

1. INTRODUCTION

In today's competitive market, commercial display manufacturers are striving to find new features to help them overtake the competition. Since consumers find Image Quality (IQ) to be one of the deciding factors when choosing a display[1], effort has been concentrated on trying to improve the image quality using various techniques. One of the bottle necks for improving IQ is the quality of the content. It has become quite common today to view video material on devices such as personal computers and mobile phones. Regardless of whether the video material is stored on the device itself or streamed from a remote server, the limitations that such devices have in storage capacity and data transfer bandwidth make it desirable to reduce the data size as much as possible by means of data compression algorithms. Unfortunately compression algorithms also introduce artifacts in the content.

It is possible to compensate for some of the artifacts caused by compression algorithms. For example, areas which have become blurred after compression can benefit from a sharpening filter. Also, the impact of blocking artifacts can be reduced by applying a blur filter[2]. On the other hand, since the visibility of each artifact can vary depending on the image content, different areas of the image can be effected more by one specific artifact than others[3,4]. Therefore, applying image enhancement filters may improve the perceived IQ in some areas of an image while making other areas worse. For example, while applying a sharpening filter will enhance areas effected by blur, it will make the blocking artifacts look worse. It is therefore important to know how the viewer evaluates the overall quality of the image if different regions of the image differ in their quality level. A more specific question is whether improving the quality of the Region Of Interest (ROI) will result in a higher IQ rating for the entire image even if the quality of some background (BG) regions have become worse in the process.

This paper describes a three-phase experiment that examines the significance of the ROI in determining the quality of the entire image. A database of images with a clear ROI was compromised to different degrees using JPEG compression. The IQ level of these images, as well as their natural ROI, was subjectively determined. The images were then manipulated to have different quality levels in the ROI and the BG regions. The overall IQ of the manipulated images was subjectively evaluated as well. This score was then compared to the subjective IQ scores of the unmanipulated images to determine whether the ROI has a stronger effect on the IQ than the rest of the image. The methodology and the experiment protocol are discussed in sections 2 and 3 respectively. Section 4 lists the results of the experiment, which are then discussed in Section 5. Finally Section 6 ends with the conclusions and mentions some possibilities for future research.

Image Quality and System Performance VII, edited by Susan P. Farnand, Frans Gaykema, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7529, 752907 · © 2010 SPIE-IS&T · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.839545

2. METHODOLOGY

2.1 Stimuli

The stimuli used in the experiment were created from 40 original images. Each image was further processed to produce 4 different versions, which resulted in a total of 160 stimuli used in the experiment. Considering the goal of the experiment, we only chose images which contained a clear ROI in the form of a face, an animal, or an object that clearly stands out from the rest of the image. Images were cropped to 600 by 600 pixels in order to have a standard size for all the images.

Each original image was degraded for the experiment with the JPEG compression function (imwrite), defined in MATLAB, using 4 different levels of compression. The compression levels used to process the images ranged between 10 (low quality) and 100 (high quality).

2.2 The eye tracker

An eye tracking system was used to determine the gaze location of the users while viewing the images. The system used in the experiment was the iView X system developed by SMI. It uses an infrared camera to track the eye movements of the user. The camera also has an infrared light mounted above the lens which is used to illuminate the eye. Since infrared falls outside the human visual spectrum, the viewer is not distracted by the light emitted from the system. In order to track the eye, the system identifies the location of the pupil and the position of the infrared light reflections from the image captured by the camera in order to calculate the gaze point. The REDIII camera used by the system has a sampling rate of 50 Hz and a tracking resolution of ± 0.1 deg. The gaze position tracking accuracy is ± 1 deg. Viewers were asked to place their head on a head rest as recommended by the eye tracking system manual in order to avoid head movements and get the highest accuracy. The head rest kept the viewer at a distance of 60cm from the screen, which represented a typical viewing distance and fell in the system's recommended operating distance of 40-60 cm. The height of the head stand was adjusted to suit the viewer and insured a comfortable and non confining seating position while performing the experiment. During all the experiments, the eye tracker was calibrated using a 13 point grid.

2.3 The experiment setup

The images were displayed on a 17-inch CRT monitor using a resolution of 1024 by 768 pixels. The experiment was controlled from a remote computer with its monitor positioned so that it was not viewable by the participant (Figure 1). In order to avoid outside elements interfering with the results, the experiment was carried out in the User-Experience Lab located in the Electrical Engineering, Mathematics and Computer Science (EEMCS) faculty building at the Delft University of Technology.

Figure 1. participants place their head on a chinrest positioned at a fixed distance from the display. The eye tracker can be seen next to the display. The experimenter controls the eye tracker and runs the experiment using another monitor which cannot be seen by the participant.

Only the experimenter and the viewer were present while performing the experiment. The lab also gave us the ability to control the light level independently of the outside lighting conditions. The illumination was kept at 70 [lux], which is a typical lighting setting in office conditions.

2.4 The participants

The experiment had a total of 75 participants. They were collected from the faculty of Computer Science at the Delft University of Technology, and were either students or staff members. It is therefore estimated that all participants possessed some experience with the type of degradation and artifacts caused by JPEG compression. When asked whether they suffered from any vision problems, they all expressed having sound (corrected) vision. This was considered sufficient to ensure that they were able to observe the difference in image quality. All participants were naive to the purpose of the experiment.

3. THE EXPERIMENT PROTOCOL

As mentioned before, the experiment included 3 separate phases. Phases 1 and 3 required people to examine images and give them a score based on their quality, while participants in phase2 were only asked to look at the images without a predefined task. The participants were divided to have 20 participants in phase1, 40 in phase2, and 15 in phase3.

The participants were informed that they would carry-out an experiment on image quality research. They were told that the position of their gaze would be recorded using an eye-tracking device. This was followed by a quick test to check whether the eye tracker locked on the participant's pupil, which was occasionally not possible due to reflections from eye glasses or to poor contrast between the pupil and the iris in the infrared spectrum. Those who passed this check were asked to start the experiment. In order to insure consistency, the instructions for the experiment were given to the participants through the computer screen, together with examples of how to perform each step. After reading all the instructions, the subjects were allowed to ask questions in order to clarify any unclear points. Once they were ready to start, the experimenter started the eye-tracker calibration process, and then started showing them the stimuli.

3.1 Phase1

Participants in phase1 were shown all 160 stimuli in the experiment. The experiment was split in 4 sessions requiring the participants to evaluate 40 images in each session. Every session contained one version of each original image presented at a certain level of compression. The system chose the image at random insuring that at the end of the session, the participant saw one version of each of the 40 original image contents in the database. In the subsequent sessions, the participant was shown one of the remaining versions of each image, which was also the case in the third and fourth sessions. The order in which the images were shown in each session was also chosen randomly by the system. This was done to avoid any systematic effect which may influence the collected data. Between the sessions, the participants were given a short break where they could take their head off the chin-rest and have something to drink. This was done to avoid strain developing in the neck and back muscles, and in order not to exhaust the eyes of the participants.

The experiment followed the single-stimulus protocol set by the ITU[5]. The participant was shown a 50% gray screen (R,G, and B values set to 127) with a white dot in the center. The participants were asked to focus their gaze on that dot while it remained on the screen for 3 seconds. The eye-tracking data collected during these three seconds was later used to refine the eye-tracker calibration. Subsequently, a randomly selected image was displayed on the screen centered on a 50% gray background. Participants were allowed to examine the image until they decided on the quality score they would like to give it. They could then use the left mouse button to go to the scoring screen, where they saw a horizontal slider bar separated into 10 equal segments with the words "Poor" on the left and "Excellent" on the right. The slider could be controlled by moving the mouse to choose the required score. Then a click on the left mouse button saved the chosen score and took the participant again to the 50% gray screen with the white dot in the center. The system then chose another image randomly from the database which was not created from the same original content as any of the previously scored images in the session. These steps were repeated until the end of the session allowing the participant to score 40 different images. After a short break, the participant started the following session by first completing the 13 point calibration step described earlier, followed by another randomly chosen 40 images, which were not shown in the previous session and each from a different original content image. This process was repeated in 2 more sessions taking each participant through the entire database of 160 stimuli.

3.2 Phase2

Here the viewers were not given any task and were only asked to view the images in a casual manner. The data collected from this phase of the experiment was later used to subjectively identify the natural ROI for the images. To avoid any deviation in the measured ROI due to a learning effect from multiple viewing of the same image content, participants only viewed one version of each image.

The second phase was performed concurrently with phase1, taking place at the same lab and using the same equipment and setup. Participants were told to simply look at the images as if they were viewing a photo album. Before the experiment started, two sample pictures were shown to the participants separated by the 50% gray screen with the white spot in the middle similar to that used in phase1. Participants were instructed to focus on the white spot while it appeared on the screen, which again gave us a uniform starting gaze position for all images and provided us with data which could be used to refine the tracker's calibration.

After completing the training, the participants went through the 13-point calibration step as before and then started viewing the images. Each image was displayed on the screen for 8 seconds followed by the 50% gray screen. Basically, each participants saw a selection of stimuli as if he completed one session of phase1. As a result, every 4 participants saw the entire set of 160 images presented at a random order, while each of them only saw one compressed version of each original content. By the end of phase2 we gathered the free looking gaze data from 10 participants for each version of the compressed images.

3.3 Phase3

The last step of the experiment was to combine different stimuli generated from the same original content but at different levels of quality. Data collected from the second phase of the experiment was used to identify the ROI of the images. In order to avoid having the size of the ROI region effect the results, only 20 of the original images, which had a similarly sized ROI, were used in phase3. The size of this area ranged from 9.5-15.8% of the entire image area (see Figure 2).

Figure 2. One of the pictures used in the experiment. All pictures is contain details both in foreground and background areas. The ROI (highlighted with green) is determined subjectively, and is chosen to occupy 9.5-15.8% of the image area.

Each stimulus used in phase3 of the experiment used two stimuli from the previous phases, where one stimulus appeared in the BG area while the other appeared in the ROI. A second stimulus was then also created from the same pair of stimuli, but with the reverse order of which stimulus appeared in the ROI and which one appeared in the BG. In total, 80 stimuli were used in phase3 which were shown in 4 sessions in a similar manner to that used in phase1.

To ensure consistency, the experiment was conducted in the same lab and under the same conditions as the two previous phases. The same scoring protocol was used as the one described in phase1 above. The eye tracker was also used to ensure uniformity in the experimental conditions, even though the data collected from the eye tracker was not needed for this phase of the experiment.

3.4 The Eye tracking data

The eye tracker collected the coordinates of the participant's gaze locations throughout each session. These data were then sorted into fixations and saccades by the eye tracking system based on the gaze dispersion within a specified amount of time. For the experiment, the system was set to consider a gaze that remained within an area of 100 pixels for 80 ms or longer to be a fixation located at the mean of the recorded coordinates. If the eye dispersion exceeded 100 pixels, the tracker indicated the movement as a saccade. So all fixations had a duration of at least 80 ms, and all saccades spanned a distance of at least 100 pixels.

While testing the SMI eye tracker system, we noticed that the recorded fixations were occasionally shifted from their correct location. This shift tended to be a constant displacement in the horizontal and vertical coordinates of the fixations for each test session. To compensate for this error in the collected data, an additional calibration step was added to the experiment. Between each two images displayed on the screen, the system displayed a 50% gray screen with a white spot in the middle for 3 seconds. The participants were instructed to keep their eyes fixed on the center of the screen (where the white spot was located). This was aimed at having a uniform starting gaze location for each participant, and eliminate any afterimage effect remaining from the previous stimulus.

Since the eye tracker recorded where the participants were looking, and we knew the coordinates of the spot that they were looking at, it was possible to use this information to compensate for the displacement in the fixation points reported by the system. The correction was performed in MATLAB by taking the mean coordinates of all fixation points collected on the gray screen for the entire session, and then applying an opposite shift to the rest of the fixation points recorded by the system.

4. RESULTS

4.1 Scoring experiment

The IQ scores collected in phases 1 and 3 of the experiment were processed using the same methodology as used in the LIVE database[6,7,8]. As a measure of how successful the experiment was, we compared the trend in the scores of phase1 to the trend of the scores found in the LIVE database. To demonstrate the level of JPEG compression that the images went through in relation to the Mean Opinion Score (MOS) they received, a scatter plot was created for both experiments, as shown in Figure 3 below.

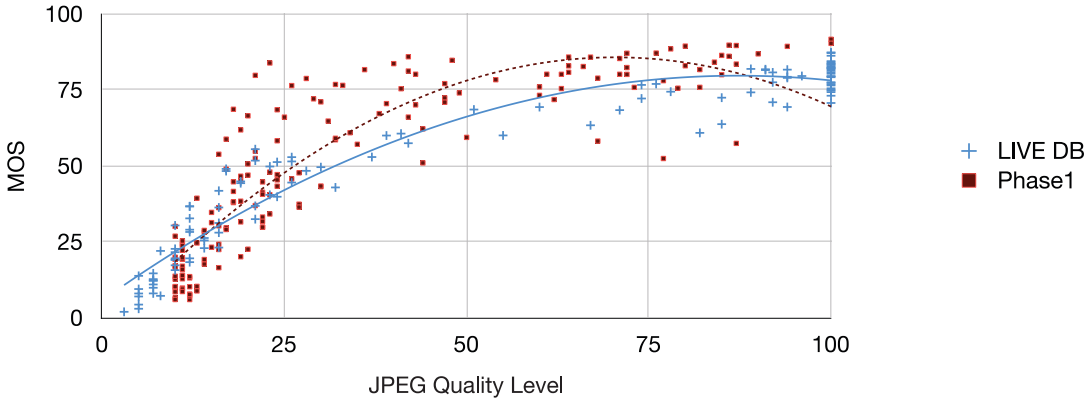

Figure 3. The relation between the JPEG compression level and the MOS given by the viewers. The figure shows the results from the LIVE database (+ symbols with a continuous trend line) and from our Phase1 experiment (squares with a dashed trend line).

From the figure, it is clear that the collected data follow a similar trend, i.e. a logarithmic increase in the MOS for images with a higher level of quality in terms of JPEG compression. There is more spread in the data collected in our experiment where some images receive a low score even though they have a high level of quality and vice versa. One possible explanation is that this can be attributed to the specific content presented with some images being very (un)sensitive to JPEG compression artifacts. With both graphs presenting a similar trend we are satisfied that the experiment matched the standard used in contemporary psychometric analysis research.

4.2 Identifying the Region Of Interest

As mentioned earlier, the images selected for this experiment were deliberately chosen to have a clearly identifiable ROI. It is expected that when observing the images without a specific task, the viewer's attention is mainly drawn towards the ROI of each image. For example, in a picture of a man standing on the beach, one would expect the head to be the region of interest in the picture, while for a picture of a face then features such as the eyes attract the viewers' attention.

The ROI for each image is determined by the eye-tracking data collected in phase2 of the experiment. Since viewers were just instructed to view images as if they were looking at a photo album, they were expected to focus on the natural ROI of the images. For each stimulus, we collected eye-tracking data from 10 different participants. The data from all participants were then combined to form 1 saliency map for each of the 160 images. Analyzing the data showed that there was no statistically significant difference between each of the 4 stimuli created from the same original content. This means that when the observers were looking at the stimuli, they were not distracted by the compression artifacts and were sine their viewing behavior did not change with the change of the compression quality. Therefore, the 4 saliency maps for each original content were combined, giving us 40 saliency maps. Eventually, the ROI was identified as the area where the top 25% of the total fixations were located.

4.3 Significance of ROI on IQ

If we assume that the ROI has the same effect on the total IQ as the BG region, then it would be possible to calculate an Expected Score (ES) for the combined stimuli used in phase3. This is possible since we have the MOS scores for all of the images used to create the combined stimuli, as well as the percentage of the area that each of these images occupy in the combined stimulus. We therefore calculate the ES of the stimuli as follows:

$$ES = (MOS_{ROI}.A_{ROI} + MOS_{BG}.A_{BG})/A_T \quad (1)$$

In other words, we assume that people evaluate the quality of the different regions and come up with an average score for the combined stimulus.

By comparing the collected MOS values to the calculated Expected-Scores, it is possible to see the effect the ROI has on the overall quality of the image. In Figure 4 we have split the data in two groups: one containing images which have a higher IQ in the ROI than in the BG (A) and one with images which have a lower IQ in the ROI than in the BG (B).

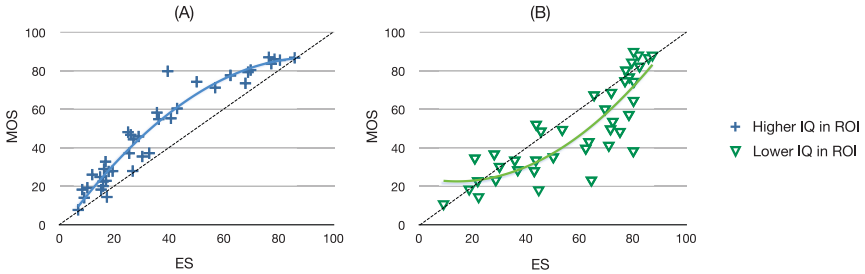

Figure 4. Comparing the calculated ES to the subjectively collected MOS. Figure 4 (A) on the left represents images which had a higher IQ level in the ROI than in the BG, while images in Figure 4 (B) on the left had lower quality in the ROI.

From the figure, it is clear that the images with higher quality in the ROI have a tendency to get a higher MOS than what the ES suggests. By looking at the trend line, one can see that this effect is at its highest for images located in the central region of the quality range. The effect diminishes when the quality of the image is too high or too low.

The effect seems to be stronger for images that have a higher quality in the ROI. Graph B shows that the effect is less prevalent for images that have the higher quality in the BG area of the image. Occasionally these images even gain an MOS that exceeds the ES.

It is also interesting to see whether the size of the quality difference between the ROI and the BG plays a role on the overall MOS of the combined stimuli. Figure 5 shows a scatter plot that attempts to illustrate this effect. In this figure, the horizontal axis represents the difference in quality between the ROI and the BG of the stimulus. The stimuli used in the experiment can fall in the negative half or the positive half of the graph depending on whether the ROI region or the BG had a higher quality. The vertical axis represents the difference between the MOS gathered in the experiment and the ES.

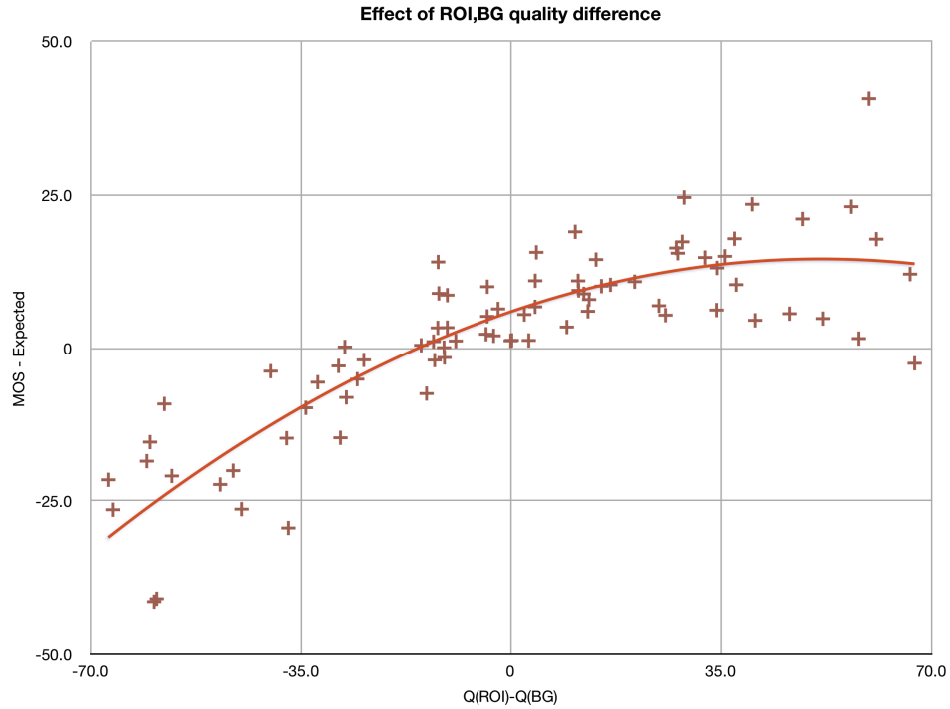

Figure 5. The horizontal axes represents the Quality level difference between the TOI and the BG regions, where the left side contains images with lower quality in the ROI than in the BG and vice versa. On the Y axes, the difference between the MOS and the calculated ES is represented. On the lower half, viewers scored the images lower than what the ES expected and vice versa.

In this graph, if there would be no difference between the effect the ROI and the BG have on the IQ, all images should lay horizontally flat on the zero value of the vertical axis, since the difference between the MOS and the ES would amount to zero. It is clear that this is not the case. Instead, values tend to be negative when the ROI has a lower quality than the BG and positive when the situation is reversed. Moreover, this effect appears to be stronger as the quality of the ROI (in relation to the quality of the BG) becomes lower. This trend becomes weaker as the quality of the background is strongly compromised in comparison to the quality of the ROI.

5. DISCUSSION

The results of the experiment clearly show that when people assess image quality, they give greater significance to some regions of the image over others. It is not possible to obtain the overall image quality by simply averaging the quality of the different regions of the image. This is shown in Figure 5 where observers gave the images a score (MOS) different

from the ones we calculated (ES) by simply averaging the quality of all image regions. As a result, the images in the figure are scattered along the vertical scale between values ranging from [-50,50] on a scale of [-100,100].

Looking at the two graphs in Figure 4 we can see that there is a relation between the ROI quality and the MOS given to a combined stimulus. Stimuli which had a lower quality in the ROI tend to score higher than expected. When the situation is reversed, the effect is not as clear. Some images with a lower quality in the ROI still gained a higher MOS. Nevertheless, the overall trend still shows that the MOS is lower than the expected ES if the quality of the ROI is lower than that of the BG area.

In order to see how this behavior is effected by the amount of quality difference between the two image regions, we can examine Figure 5 again. The horizontal axis represents the difference in quality between the ROI and BG regions. Looking at the lower left corner of the graph, one can see that as the quality of the ROI gets more degraded, the MOS shifts further away below the ES. In the center of the figure, where the quality level in both image regions is very close, we see that the MOS and the ES are very close as well. One can notice that the MOS is slightly higher than the ES. This shift in the results can be attributed to the inaccuracy of the scoring experiment used to collect the MOS in the first and third phases of the experiment. As the quality of the ROI continues to increases (towards the right side of the graph), the difference between the MOS and the ES stops growing and starts to diminish at the extreme end of the graph. One can therefore conclude that even if the degradation is only present at the BG region, at a certain point the degradation becomes so bad that it plays a bigger role in determining the MOS for the entire image.

6. CONCLUSIONS

From the above discussion, we see that it is important to take the ROI of the image into consideration when trying to apply any manipulation to the original image content. As shown in Figure 5, if the manipulation lowers the quality of the ROI, then the perceived IQ of the entire image will be lower even if the majority (84.2% to 90.5%) of the image area has benefited from the manipulation. It is therefore risky to apply naive image enhancement algorithms which do not take the ROI into consideration.

When the quality of the ROI is higher than that of the BG, the viewers tend to give the image a higher quality score than its average quality level (as shown in Figure 4 (A)). Though as the top right corner of Figure 5 shows, the relation is not as clearly visible when the ROI's quality is lower than the BG quality. Furthermore, if the IQ of the BG area is highly compromised (at the very end of the right side of the figure) we see that the overall image quality starts to drop even though the quality of the ROI is very high.

It would be interesting to extend this study to video content. Since the dynamic nature of video lends greater significance to the ROI, we would expect the results to be more pronounced. It is also important to see whether these findings will hold true for other types of image manipulations than JPEG compression. Since the human eye is not good in detecting details in the periphery, it is possible that the observers are not capable of detecting the lower (or higher) quality in the BG region. This is not the case for differences in luminance levels, where the human eye has a better performance on the entire field of vision. As a result, a similar experiment using manipulations in brightness or contrast can lead to totally different results. We therefore feel that there is still much to explore in regard to the significance of the ROI in determining IQ and plan to do more experiments on this topic in the future.

REFERENCES

1. Engeldrum P., [Psychometric Scaling], Imcotek Press, Winchester, Massachusetts, USA, (2000).
2. Reeves H. C. and Lim J. S., “Reduction of blocking effects in image coding”, Optical Engineering, vol. 23(1), 34–37 (1984).
3. Girod B., “The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals”, Proc. of the SPIE Human Vision, Visual Processing, and Digital Display, vol. 1077, 178–187 (1989).
4. Liu H. and Heynderickx I., “A Simplified Human Vision Model Applied to a Blocking Artifact Metric,” Proc. CAIP, LNCS 4673, 334-341 (2007).
5. Recommendation BT.500-10, “Methodology for the subjective assessment of the quality of television pictures”, ITU-R (2000).

6. Sheikh H.R., Wang Z., Cormack L. and Bovik A.C., "LIVE Image Quality Assessment Database Release 2", http://live.ece.utexas.edu/research/quality.
7. Sheikh H.R., Sabir M.F. and Bovik A.C., "A statistical evaluation of recent full reference image quality assessment algorithms", IEEE Transactions on Image Processing, vol. 15, no. 11, 3440-3451 (2006).
8. Wang Z., Bovik A.C., Sheikh H.R. and Simoncelli E.P., "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing , vol. 13, no. 4, 600- 612 (2004).