Studying the effect of optimizing image quality in salient regions at the expense of background content

Hani Alers
Judith Redi
Delft University of Technology
Man-Machine Interaction
Delft 2628 CD, The Netherlands
E-mail: hani.alers@gmail.com

Hantao Liu
University of Hull
Department of Computer
Hull HU6 7RX, United Kingdom

Ingrid Heynderickx
Delft University of Technology
Man-Machine Interaction
Delft 2628 CD, The Netherlands
and
Philips Research Laboratories
Group Visual Experiences
Eindhoven 5656 AA, The Netherlands

Abstract. Manufacturers of commercial display devices continuously try to improve the perceived image quality of their products. By applying postprocessing techniques on the incoming signal, they aim to enhance the quality level perceived by the viewer. These post-processing techniques are usually applied globally over the whole image but may cause side effects, the visibility and annoyance of which differ with local content characteristics. To better understand and utilize this, a three-phase experiment was conducted where observers were asked to score images that had different levels of quality in their regions of interest and in the background areas. The results show that the region of interest 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. Based on the subjective data we propose a model to predict the overall quality of images with different quality levels in different regions. This model, which is constructed on empirical bases, can help craft weighted objective metrics that can better approximate subjective quality scores.

1 Introduction

In today’s competitive market, commercial display manufacturers are striving to find new features to help them overtake competition. Since consumers find image quality (IQ) one of the deciding factors when choosing a display, research and development effort has been concentrated on improving IQ using various techniques. For some applications, however, the quality of the content is one of the bottlenecks. 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 video 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 that have become blurred after compression can benefit from a sharpening filter. On the contrary, the impact of blocking artifacts may be reduced by applying a blur filter. Since the visibility of each artifact may vary depending on the image content, one part of an image may be affected more by a specific artifact than other parts. Therefore, applying an image enhancement filter may improve the perceived IQ in some areas of an image while making other areas worse. For example, applying a sharpening filter will enhance areas affected by blur, while it will make blocking artifacts more visible. 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) results in a higher IQ rating for the entire image even if the quality of some background (BG) regions has become worse.

The work discussed in this article shows that subjectively measured mean opinion score (MOS) is different from an estimated score obtained by averaging the quality of all
image regions. Nonetheless, the latter is what most quality estimation algorithms do; they locally estimate (based on pixel values) a quality score and then average these scores over the entire image. As such, the calculated overall quality is an (area weighted) average of the local quality in different regions of the image. More advanced quality estimation algorithms include saliency weighting; i.e., the local quality values are weighted with the local saliency, as such giving more value to quality in the ROI than to quality in the BG. So far, however, the weighting strategy for adding saliency has not been determined. Attempts to determine this weighting strategy have largely depended on trial and error. As such, this paper contributes in quantifying the optimal weighting strategy.

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 in quality to different degrees using JPEG compression. The IQ levels of these images, as well as their natural ROI, were subjectively determined with the help of an eye-tracking system. 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. These scores were then compared to the subjective IQ scores of the nonmanipulated images to determine whether the ROI had a stronger effect on IQ than the rest of the image.

The methodology and the experimental protocol are discussed in Secs. 2 and 3, respectively. Section 4 lists the results of the experiment, which are then discussed in Sec. 5. Finally, Sec. 6 summarizes the conclusions and mentions some possibilities for future research.

2 Experimental Setup

2.1 Stimuli

The stimuli used in the experiment were created from 40 original images. All were natural images containing humans, animals, or structures. Considering the goal of the experiment, we chose only images that contained a clear ROI in the form of a face, an animal, or an object that clearly stood out from the rest of the image. Images were cropped to 600 by 600 pixels (corresponding to a viewing angle of 20.2 deg) in order to have a standard size for all images.

Each image was further processed to produce four different versions, which resulted in a total of 160 stimuli used in the experiment. These versions were created with the JPEG compression function (imwrite), defined in MATLAB®. The compression parameter for the (imwrite) function in the four compression levels used to process the images ranged between 10 (low quality) and 100 (high quality), and were different for the 40 different originals. Some example images are shown in Fig. 1.

2.2 Eye Tracker

To record the gaze location of the users viewing the images, an eye-tracking system (i.e., iView X system developed by SMI) was adopted. It tracks the eye movements of the users with an infrared camera, recording the reflections of a small infrared source at the eye’s retina. Since infrared falls outside the spectral range of sensitivity of the human visual system, the viewers were not distracted by the infrared light emitted by the eye tracker. The REDIII camera used by the eye tracker had a sampling rate of 50 Hz and a tracking resolution of ±0.1 deg. Viewers were asked to place their head on a head rest as recommended by the system’s manual. The head rest restrained head movements and kept the viewer at a distance of 60 cm from the screen, which represented a typical viewing distance and fell in the system’s recommended operating distance of 40 to 60 cm. The eye tracker was calibrated using a 13-points grid and resulted in a gaze position tracking accuracy of ±1 deg. The height of the head stand was adjusted to suit the viewer and a comfortable and nonconfining seating position was insured while performing the experiment.

2.3 Facilities

The experiment was carried out in an isolated room. Only the experimenter and participant were present during the experiment. All stimuli were displayed on a CRT monitor with a resolution of 1024 by 768 pixels and an active screen area of 365 × 275 mm. The experiment was controlled from a remote computer with its monitor positioned so that its content was not visible to the participant to avoid distractions (see Fig. 2).
2.4 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 differences in image quality.

The participants were informed that they would carry out an experiment on image quality research. They were told that their eye movements would be recorded using an eye-tracking device. However, they were not informed about the goal of the experiment or how the data were going to be analyzed in order to not reduce the influence on their viewing behavior. After they gave their consent, a quick test was performed to check whether the eye tracker locked on the participant’s pupil. The latter was occasionally not possible due to reflections from eye glasses or due to poor contrast between the pupil and the iris in the infrared spectrum. These participants had to be excluded from the experiment and were replaced by new ones.

3 Experimental Protocol

As mentioned before, the experiment included three separate phases. Phases 1 and 3 required people to examine images and to give each image a score based on the perceived quality. Participants in phase 2 were only asked to look at the images without a predefined task. The participants were divided such that we had 20 participants in phase 1, 40 in phase 2, and 15 in phase 3. A larger number of participants was needed in phase 2 to identify the natural ROI of the images (see Sec. 4.1). Phase 3 required fewer participants since the eye-tracking data were not going to be analyzed and the number of images was lower than that of phase 1. Each phase of the experiment adopted a within-subject design, where changes in the dependent variable (that is, IQ in phases 1 and 3 and visual attention (VA) deployment in phase 2) are analyzed as they appear across different images rather than across groups of different test subjects. However, different participants were recruited for each phase.

Participants who passed the check with the eye tracker were asked to start the experiment. In order to ensure 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 having read all instructions, the subjects were allowed to ask clarifying questions. Once they were ready to start, the experimenter started the eye-tracking calibration process and then started showing them the stimuli. To avoid introducing a bias in the results, each participant saw the corresponding stimuli in a different random order.

3.1 Phase 1

Participants in phase 1 were shown all 160 stimuli of the experiment. The experiment was split in four 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 images in the database. In the subsequent sessions, the participant was shown one of the remaining versions of each image. The order in which the images were shown in each session was chosen randomly by the system. 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 International Telecommunication Union.29 The participants were shown a 50% gray screen (R, G, and B values set to 127) with a white dot in the center. They were asked to focus their gaze on that dot while it remained on the screen for 3 s. The eye-tracking data collected during these 3 s were later used to refine the eye tracker’s calibration (see also Sec. 3.4). Subsequently, a randomly selected image was displayed on the screen centered on a 50% gray BG. Participants were allowed to examine the image until they decided on the quality score. 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. These steps were repeated until the end of the session, in which the participants had to score 40 different images. After a short break, the participants started the following session by first completing the 13-point calibration step described earlier, followed by another 40 images randomly chosen. This process was repeated in two more sessions taking each participant through the entire database of 160 stimuli.

3.2 Phase 2

In this phase, 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 were later used to subjectively identify the natural ROI of the images. To avoid any deviation in the measured saliency due to a learning effect from viewing the same image content multiple times, participants viewed only one (compressed) version of each original image.

The second phase was performed concurrently with phase 1, 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 images were shown to the participants. These images were separated by the 50% gray screen with the white dot in the center, similarly as in phase 1. Participants were instructed to focus on the white dot while it appeared on the screen, which again gave us a uniform starting gaze position for all images and provided us with data that could be used to refine the eye tracker’s calibration.

After completing the training, the participants went through the 13-point calibration step as before and then started viewing the stimuli. Each stimulus was displayed on the screen for 8 s followed by the 50% gray screen. Basically, each participant saw a selection of all 160 stimuli...
as if he completed just one session of phase 1. As a result, every four participants saw the entire set of 160 images presented at a random order. As a consequence, by the end of phase 2, we gathered the free looking gaze data of 10 participants for each compressed version of the 40 original images.

### 3.3 Phase 3

The last step of the experiment used stimuli generated from the same original content, but with a different level of quality for the ROI and BG. Data collected from phase 2 of the experiment were used to identify the ROI of the images. In order to avoid the size of the ROI region affecting the results, only 20 of the original images with a similarly sized ROI were used in phase 3. The size of the ROI ranged from 10 to 16% of the entire image area corresponding to a viewing angle of 2.0 to 3.2 deg (see Fig. 2).

From the original 160 images used in this experiment, only 80 were chosen for phase 3 (20 of the 40 original images), which had the most clear and uniquely identifiable ROI. Each stimulus in phase 3 contained data from two stimuli of phase 1. Basically, from every two versions of an image with different levels of quality in phase 1, the contents of the ROI were swapped between the two images. This created two combined images with different levels of quality inside and outside the ROI (see Fig. 3). The edge between the two regions was softened using a $3 \times 3$ pixel Gaussian function. In total, 80 stimuli were used in phase 3. They were each shown to participate in four sessions in a similar manner as used in phase 1. Figure 4 shows an example of the resulting combined images.

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 as the one described in phase 1 was used. The eye tracker was also used to ensure uniformity in the experimental conditions to make sure any change in the data is not caused by not using the eye tracker in phase 3. The data collected from the eye tracker were not needed for this phase of the experiment.

### 3.4 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 (viewing angle of 3.4 deg) for 80 ms or longer as a fixation. Its location was calculated as the mean of the coordinates over the entire length of the fixation. 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 eye tracker, we noticed that the recorded fixations were occasionally shifted from their correct location. This shift tended to be a constant displacement in horizontal and vertical direction of all fixations in a 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 dot in the center. The participants were instructed to keep their eyes fixed on the dot. As such, we aimed at having a uniform starting gaze location for each participant. Since the eye tracker recorded where the participants were looking, and we knew the coordinates of the dot that they were supposed to look at, it was possible to compensate for the possible shifts in fixations. 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.

The iView X eye-tracking system mainly collects fixations that need to be converted into saliency maps. These maps show a visual representation of the probability that a location of the scene is attended by the average observer. To create the saliency map for a given image $i$, first a fixation map is created that includes all the fixation locations from all observers recorded for that image. These fixation maps are then converted to saliency maps by applying a Gaussian patch with a width $\sigma$ to each fixation in the map. The width $\sigma$ of the Gaussian patch is chosen to be 2 deg of visual angle, approximating the size of the fovea of the human eye and sufficiently accounting for inaccuracy in the measurement.
of the fixations. A mean saliency map that takes into account all fixations of all subjects is then calculated as follows:

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

where $S_i(k, l)$ indicates the saliency map for stimulus $i$, $(k, l)$ refers to a pixel in the saliency map of size $M \times N$ pixels (i.e., $ke[1, M]$ and $le[1, N]$), $(x_j, y_j)$ indicates the spatial coordinates of the $j$th fixation ($j = 1 \ldots T$), $T$ is the total number of all fixations over all subjects, and $\sigma$ indicates the standard deviation of the Gaussian. The intensity of the resulting saliency map is linearly normalized to the range [0, 1].

### 4 Results

#### 4.1 Natural 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 toward the natural ROI of each image. For example, in a picture of a man standing on the beach, one would expect his head to be the ROI of the picture, while for a picture of a face, features such as the eyes usually attract viewers’ attention.

The ROI for each image was determined by the eye-tracking data collected in phase 2 of the experiment. For each of the 160 stimuli, the data of all 10 participants were averaged into one saliency map. Since the viewers saw compressed versions of the images in phase 2, we needed to determine whether the quality level affected their viewing behavior. To that effect, the highest-and lowest-quality version of each image were distributed into two separate sets. Then an independent sample $t$-test was performed to determine whether the difference in quality resulted in a difference in the viewing behavior with the independent variable being the quality level (high or low) and the dependent variable being the similarity score that shows how similar the saliency maps are to each other. The test showed no significant difference in the similarity of the saliency maps between the high- and low-quality images ($p = 0.48, F = 5.94$). This shows that when the observers were looking at the stimuli, they were not distracted by the compression artifacts and their viewing behavior did not change with the change in compression level. Therefore, the four saliency maps for each original content were averaged, giving us 40 saliency maps.

#### 4.2 Defining the Region of Interest

The ROI for each image is extracted from the saliency maps. These maps were normalized on a scale of 0-1 representing the level of intensity of the saliency heat-maps, the ROI was identified as the area with the top 25% of the range (i.e., scoring 0.75 to 1 on the heat-map scale).

#### 4.3 Significance of ROI on IQ

The IQ scores collected in phases 1 and 3 of the experiment were processed to calculate the MOS. First the $Z$-scores where calculated for each image. Then the standard normal distribution of the resulting $Z$-scores was taken to give a score in the range (0 to 1). From the MOS obtained in phase 1, it is possible to estimate the expected score (ES) of a combined stimulus in phase 3, assuming that the observer will average out the overall quality of the image without giving more importance to the quality of a specific region. In that case, the ES is the weighted sum of the MOS of each stimulus, as obtained in phase 1, weighted only with the percentage of area of the ROI and BG, respectively. Hence, under this assumption, we would calculate the ES as follows:

$$ES = MOS_{ROI} \cdot A_{ROI} + MOS_{BG} \cdot A_{BG}. \quad (2)$$

with $MOS_{ROI}$ and $MOS_{BG}$ being the scores of the images used in the ROI and in the BG regions, respectively, and $A_{ROI}$ and $A_{BG}$ are the ratios of the area of the ROI and the BG regions, respectively, to the entire image. The way the ES is calculated is also illustrated in Fig. 3. By comparing the collected MOS values of all stimuli of phase 3 to their estimated ES, it is possible to extract the effect the ROI has on the overall quality of an image. Figure 5 presents these data split up in two data groups: one with images that have a higher IQ in the ROI than in the BG (a) and one with images that have a lower IQ in the ROI than in the BG (b). Figure 5(a) clearly shows that the images with higher quality in the ROI have a tendency to get a higher MOS than what would be expected from the ES. Looking at the trend line, this effect seems to be stronger 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. A similar tendency is seen in Fig. 5(b); images with a lower quality in the ROI tend to get a lower MOS than what would be expected from the ES. The latter effect, however, seems to be weaker than the one found for images with a higher quality in the ROI. Figure 5(b) shows that the effect is less prevalent for images that have a 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 stimulus. Figure 6 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. All stimuli fall either 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 collected from phase 3 and the ES. If the effect of the ROI and BG on overall IQ was equal, all data points in Fig. 6 would lie horizontally on the $Y = 0$ axis, since the difference between the MOS and the ES would then be zero for all stimuli. 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 difference in quality between the ROI and BG increases, the latter being especially the case for images with a higher quality in the BG than in the ROI (so, at the negative side of the $X$ axis in Fig. 6). This trend is weaker when the quality of the BG is strongly compromised in comparison to the quality of the ROI.
4.4 Modeling the Influence of ROI on Overall IQ

Using the data collected from the experiment, it is possible to estimate how much more important the ROI is in determining the overall quality of an image. To do that, we again look at Eq. (2) used to calculate the ES. Since we now know that there is a difference in how much each region affects the overall perceived quality, we calculated a more accurate weighted expected score (WES) by introducing two weighting parameters to the equation, resulting in

$$WES = MOS_{ROI} \cdot A_{ROI} \cdot w_{ROI} + MOS_{BG} \cdot A_{BG} \cdot w_{BG}.$$  

(3)

where \(w_{ROI}\) determines the weight of the ROI and \(w_{BG}\) the weight of the BG region on the overall perceived quality. To calculate the values of these weights, a linear regression analysis was performed. The analysis used the MOS of the combined images as the dependent variable and the quality parameter for each region multiplied by its corresponding area as the independent variables. The analysis returned the values \(w_{ROI} = 3.80\) (\(p < 0.001\), 95% confidence interval 3.34 to 4.27), and \(w_{BG} = 0.65\) (\(p < 0.001\), 95% confidence interval 0.58 to 0.71). The overall model fit had an \(R^2 = 0.975\). The resulting relation is depicted in Fig. 7, again for the two groups of data separately, i.e., in Fig. 7(a) for the images with a higher IQ in the ROI than in the BG and in Fig. 7(b) for the images with a lower IQ in the ROI than in the BG.

To test the stability of this fit, the 80 stimuli of experimental phase 3 were split into two subgroups of 40 stimuli each. Both subgroups spanned the entire range of the quality scale. The two counterparts of each combined picture (i.e., one with a higher IQ in the ROI and the other with a higher IQ in the BG) were joined in the same subgroup in order to avoid having the same image content repeated in both subgroups and thereby influencing the analysis. We then conducted a linear regression analysis in the same manner as described above on one of the two subgroups. This analysis yielded the values \(w_{ROI}^{sg} = 3.54\) (\(p < 0.001\), 95% confidence interval 2.85 to 4.25) and \(w_{BG}^{sg} = 0.68\) (\(p < 0.001\), 95% confidence interval 0.58 to 0.78). The overall model fit had an \(R^2 = 0.974\).

Both weighting factors were close in value to the ones found for the whole ensemble of stimuli, indicating that the result of the fit was not very sensitive to the particular selection of stimuli used. Subsequently, the new values of \(w_{ROI}^{sg}\) and \(w_{BG}^{sg}\) were used in Eq. (3) to calculate the WES of the second subgroup of stimuli (see Fig. 8).

Finally, a similar plot as the one shown in Fig. 6 is generated using the WES scores from the second subgroup of stimuli and the result is shown in Fig. 9. The data points are now scattered around the \(Y = 0\) axis, indicating that with the proper weighting factors for the quality of the ROI and the BG, the overall quality of an image, locally varying in quality, can be predicted. By examining the values of \(w_{ROI}^{sg}\) and \(w_{BG}^{sg}\), we can conclude that the quality of the ROI is about five times more important than the quality of the BG.

![Fig. 6](image_url) The horizontal axis represents the difference in quality between the ROI and the BG regions, with the left side containing images with lower quality in the ROI than in the BG, and right side containing images with higher quality in the ROI than in the BG. The vertical axis, the difference between the MOS and the calculated ES is given. On the lower half, viewers scored the images lower than what would be expected from the ES, while on the upper half, viewers scored the images higher than what would be expected from the ES.

![Fig. 5](image_url) Comparing the calculated ES to the subjectively collected MOS. (a) Images that had a higher image quality (IQ) level in the ROI than in the BG. (b) Images that had a lower quality in the ROI than in the BG.
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. The subjectively measured MOS is clearly different from the estimated score obtained by averaging the quality of all image regions, even when taking into account their relative area in the image. Results of phase 3 of the experiment demonstrated that there is a relation between the ROI quality and the MOS given to a combined stimulus. Stimuli that had a higher quality in the ROI mostly scored higher than expected. Stimuli with a lower IQ in the ROI scored lower than expected, though this effect was less clear. The extent to which the subjectively determined MOS differed from the ES was affected by the amount of quality difference between the two image regions, as was shown in Fig. 6. Looking at the lower left corner of the figure, 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, the MOS and the ES are very close as well. One can notice, however, that the MOS is slightly higher than the ES, where equality is expected. The shift may be attributed to differences in scoring between the first and third phases of the experiment, for example, as a consequence of the different groups of participants used. As such, this shift may be considered as an estimation of the reproducibility of the quality scores over the whole experiment. Since this shift is considerably smaller than the difference between the MOS and ES measured at both ends of the quality (difference) range, we are convinced that the impact of the quality of the ROI on the overall quality is not an artifact of the limited reproducibility of the quality scores. As the quality of the ROI continues to increase (toward the right side of the graph), the difference between the MOS and the ES stops growing and even seems to diminish at the extreme end of the graph. This seems to suggest that even if the degradation is only present at the ROI region, at a certain point, the degradation becomes so bad that it

![Fig. 7 Comparing the calculated weighted expected score (WES) to the subjectively collected MOS. (a) Images that had a higher IQ level in the ROI than in the BG. (b) Images that had a lower quality in the ROI than in the BG.](image1)

![Fig. 8 Comparing the calculated WES to the subjectively collected MOS for the stimuli belonging to the second subgroup. (a) Images that had a higher IQ level in the ROI than in the BG. (b) Images that had a lower quality in the ROI than in the BG.](image2)
plays a bigger role in determining the MOS for the entire image.

We also quantified the effect of the quality of the ROI on the overall quality using linear regression. The resulting values, i.e., $w_{ROI} = 3.803$ and $w_{BG} = 0.648$, suggest that the ROI region is more than five times more significant in determining the overall quality of an image than the BG region. This is even more impressive when one takes into account that the ROI in the used images occupied only 10 to 16% of the entire image area. Subsequent analysis also proved that this simple linear regression model already resulted in a considerable improvement in predicting the MOS value of stimuli with a different quality level in different areas of the image.

At this point, one can wonder whether changing how the ROI is defined will influence how much more significant it will be in determining the overall image quality. Since we used in our experiment only images with a clear ROI, which was always occupied by a human or an animal, the ROI was well defined. One should keep in mind that during the experiment, we used a white dot in the intermediate screen between stimuli and asked the participants to focus on this dot. This procedure helped us to refine the calibration of the eye tracker, but may have introduced a center bias in the saliency map as well. The effect of a fixed starting point on a center bias in a saliency map may have occurred in the first fixations, but is expected to rapidly disappear over time (as extensively discussed in the literature). Indeed, the example given in Fig. 3 clearly illustrates that the saliency map calculated over the full presentation time of 8 s results in the expected ROI, away from the image center. Hence, we are convinced that we were able to accurately find the ROI in all images of our data set. In real practice, however, not all images have a clear ROI, and the ROI has to be determined with an algorithm. Both aspects will reduce the accuracy with which the ROI is defined, and most probably the resulting ROI will be more scattered than what we used to establish our model. As a consequence, the importance of the quality of the ROI may be overestimated in real applications. Nonetheless, most images taken for entertainment purposes contain a main subject that constitutes an ROI, and the performance of ROI estimation algorithms is expected to further improve. Thus, the quality of the ROI being five times more important than the quality of the BG is a ratio that may be used for a wide variety of images uploaded online today. One can even consider an alternative approach in which the importance factor is decreased when the area of the ROI becomes bigger, but this approach would need further research to establish the relationship.

A limitation of our study is that all images in the database were degraded using JPEG compression. JPEG is still one of the de-facto formats used to save digital imagery, but its specific nature may have affected the low importance of the quality of the BG to the overall quality score. 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 as a consequence of blockiness or blur in the BG region. It is, however, good to realize that most postprocessing manipulations address spatially detailed information, so the relative importance of the quality of ROI and BG is expected to hold for a broad range of signal processing algorithms. An exception may be artifacts with a temporal nature, since our peripheral vision is more sensitive to temporal artifacts than our foveal vision. Hence, temporal artifacts in the BG may be more easily detected or may be more annoying than temporal artifacts in the ROI, so these artifacts may change the relative importance in overall quality of ROI and BG.

Finally, it is good to realize that there are already several algorithms and image formats available that can encode different regions of an image at different quality levels. There are also mechanisms available that can objectively estimate the ROI. It is therefore already possible to implement the functionality that would optimize the encryption of different regions of images while predicting how it will affect the overall quality. This can be practical for saving content on mobile devices where memory-space is limited, or when making content for the web to save bandwidth.

6 Conclusions

Our results have proven that it is important to take the ROI of an image into consideration when trying to apply any manipulation to original image content with the aim of improving its overall IQ. 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 to 90%) of the image area has benefited from the manipulation. It is therefore risky to apply naive image enhancement algorithms that 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. We propose a simple model to estimate the overall perceived quality from the different quality levels of ROI and BG regions. This model illustrates that the quality of the ROI is about five times more important for the overall quality judgment than the quality of the BG.

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. On the other hand, video is expected to be
more prone to temporal artifacts, which may be more annoying in the BG (our peripheral vision) than in the ROI (our foveal vision). Further research is needed to establish which of the two video-related aspects dominate the relative importance of quality in the ROI and BG.

References

Judith Redi is an assistant professor in the Man-Machine Interaction group of Delft University of Technology, where her main research interests are visual perception, image quality study and modeling, intelligent lighting applications.

Hantao Liu received his MSc degree from the University of Edinburgh, Edinburgh, United Kingdom, and his PhD degree from Delft University of Technology, Delft, The Netherlands, in 2005 and 2011, respectively. He is currently an assistant professor (lecturer) with the Department of Computer Science at the University of Hull, Hull, United Kingdom, leading research on computational perception and image computing. His research interests include image analysis, multimedia signal processing, and human visual perception.

Ingrid Heynderickx joined Philips Research in 1987 and since then has been involved in various research areas: optimization of the optics of LCDs, characterization of the rheological behavior of liquid crystalline polymers, and the functional evaluation of personal care devices.