Explaining multivariate image quality - Interpretation-Based Quality Approach

Jenni Radun, Toni Virtanen, and Göte Nyman, University of Helsinki; and Jean-Luc Olives, Audio-Visual Entity, Nokia-Group (Finland)

Abstract

We introduce an Interpretation-Based Quality (IBQ) estimation method for measuring subjective image quality. IBQ method reveals essential quality dimensions of test images, and provides informative descriptions related to their subjective, perceived properties. This makes it especially suitable for examining multivariate and high-level image quality. We describe the use of the method in measuring the subjective image quality of digital cameras. IBQ measurement data was obtained from naïve subjects (N=29). The subjective estimations of camera performance were evaluated in the test using pictures of a studio scene taken in three different light conditions and presented as paper photographs. The subjects estimated the overall quality of each image and described the most distinctive features of its image quality. Together these measures and descriptions offer a quantitative and qualitative insight into e.g. what is perceived as pleasing or disturbing in the image quality of the studied camera.

Introduction

Subjective image quality measurement schemes rely typically on standard psychophysical approach to perception, which aims at determining relevant thresholds for detecting and discriminating certain features in the test stimuli. Typically, the effects of a single characteristic/attribute and its contribution to image quality are evaluated. However, even a change in a single feature of an image can alter the conveyed visual message in several ways. Especially, when several image quality features change simultaneously their combined effect to subjective perceptual experience becomes difficult to evaluate. In practice, however, this is just the type of situation the users meet in using their imaging devices.

One way to estimate the quality of pictures, in which multiple variables change, is to use a single measure of overall image quality or of the amount of impairment perceived in the picture. One this kind of measure, aimed especially at estimating image compression, is anchored Mean Opinion Score (MOS) [1]. These kinds of measures are good and fast in determining the differences in image quality, but they do not tell how the quality varies. The more detailed description of image quality is often done with attribute scaling. The attribute scales usually aim the attention to technical details, because they are often constructed by image quality experts. However, the users of imaging devices are typically naïve in relation to these technical defects. Hence, for example Engeldrum [2] recommends the use of naïve observers when simulating a response of a typical customer for a certain imaging device. Furthermore, expert and naïve observers may evaluate image quality differently: expert observers tend to focus more on technical defects related to certain technology than the naïve observers do [3]. Therefore, we claim that to complement overall estimation of image quality with further description of image quality features, naïve observers must have a possibility to use their own criteria.

Here we introduce a simple way to use the Interpretation-Based Quality (IBQ) method in comparing subjective image quality of digital cameras. The method provides data of the subjective differences seen in the pictures produced by different cameras in a free description situation. This corresponds to the situation that the customers are faced with, when evaluating imaging devices. We combine anchored Mean Opinion Score (MOS) estimation [1] and free descriptions of most important characteristics seen in image quality, in order to get a full view of what subjective properties of the images appear to change when image parameters or the camera are changed. The MOS tells only that the quality changes, but free descriptions should help to explain these changes. The performance of different cameras is tested by taking a picture of a studio scene in three different light conditions. This gives an insight of how the cameras perform with respect to other cameras in different imaging conditions.

The main questions in this study are:

1. Are naïve observers capable and reliable in describing the subjective image quality of different images using their own words and without training?
2. Can free descriptions of image quality help to explain camera performance?

Methods

In the free description method the subjects defined the most important characteristic of the image quality, from their own perspective - without any guidance to observe or interpret certain predetermined aspects of image quality. In addition to this, they also gave numerical estimations of the goodness of image quality (MOS) of each picture on an 11-point scale (0 - 10).

Material

Pictures of a studio scene were used as test material. The scene has been developed to show different image quality artifacts. Altogether 29 digital cameras were included in testing (table 1 presents the megapixels of cameras). Pictures were taken in three different luminance levels, using D65 to produce 1000 lx and halogen sources for 100 lx and 10 lx. Thus the three light conditions were D65 1000, HALO 100, HALO 10. Light condition D65 1000 corresponds to a cloudy day outside, HALO 100 to normal lighting conditions indoors and HALO 10 has a very low level of light similar to, for example, candle light. Images were scaled to match the photo printer resolution, transformed to the printer color space by using ICC profiling. Pictures were printed on paper and had the size of 10x13 cm. Altogether, 95 images were used of which six were control (replica) images to estimate the consistency of observers' evaluations.

Table 1. Table 1 shows the number of cameras with different megapixels in the experiment.

Number of cameras	Megapixels
10	0.31
7	1.3
7	2
4	3
1	6.3
29	Total

Viewing conditions

The lighting in the test laboratory was 500 lx and 6000 K on the image surface. The viewing distance was ~40 cm.

Observers

Altogether 30 Finnish-speaking adults participated. Before the participation, observers passed vision tests for near visual acuity, near contrast sensitivity, and colour vision. One observer was excluded, because of inconsistent quality estimations given to control pictures, so the final number of observers in the study was 29 of which 20 were women and 9 were men.

Procedure

The observers estimated the image quality of each picture on an 11-point scale (0=poor quality, 5=moderate quality, 10=excellent quality).The pictures were presented to each observer in random order. During the whole procedure observers had examples of a high quality picture (representing suggested quality estimation of 10) and low quality picture (representing suggested image quality estimation of 1) (see [1]). Using their own words, observers also wrote down the most distinctive characteristic of image quality for each picture.

Analysis

The observers' free description answers were summarized with more general concepts (codes) that represent the attributes. This coding was done in a program Atlas.ti (Scientific Software Development, Berlin, Germany) according to the grounded theory principles [4], where the coded concepts are taken straight from the data, not from the researchers' hypotheses. This summarization makes it possible to analyze the data statistically, which was done in the program SPSS (SPSS, Inc., Chicago, Illinois, USA).

Results

The quality estimations were different for cameras ($F(9,209)=109.6$, $p<0.001$), light conditions ($F(2,46)=441.9$, $p<0.001$) and there was an interaction of cameras and light conditions ($F(13,302)=26.1$, $p<0.001$) (two-way repeated ANOVA). The MOS data in Figure 1 shows how the overall image quality of different cameras changes in different lighting conditions.

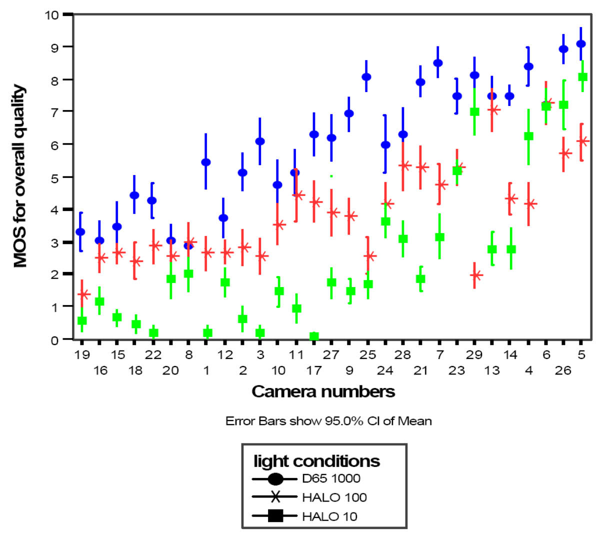
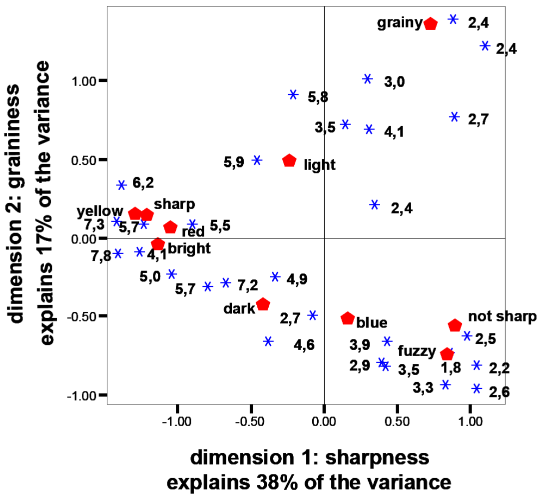


Figure 1. *The image quality estimation means (MOS) for the cameras in different light conditions and their 95% confidence limits show that the observers estimated the image quality of pictures in the similar way.*

To complement this data and to find the underlying subjective effects of these systematic quality variations, the second task was accomplished, in which we asked the observers to tell what was the distinctive characteristic in each picture's image quality. After coding these descriptions, 10 most used descriptions were selected for further analysis. We applied correspondence analysis to this data using as input the cameras and the 10 attributes used to describe the pictures produced by a certain camera in different lighting conditions. This analysis reveals the similarities of cameras and their relation to the subjective attributes in a relative space (Figure 2). The three significant dimensions obtained in the analysis explain 70% of the variance (inertia >0.2). The first dimension can be interpreted as describing the image quality in terms of sharpness and brightness. The second dimension distinguishes cameras with grainy images from those being "not sharp" or "fuzzy". The third dimension separates cameras with extremely dark images and those that have overall colour red. The halogen light explains the redness and the low illumination level explains the darkness of pictures.

2a



2b

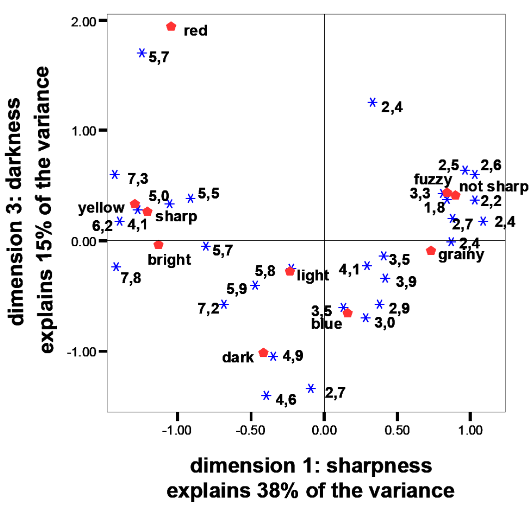


Figure 2 a & b. *In correspondence analysis with cameras and the attributes used to describe the pictures taken with a certain camera, we got three significant dimensions. The red dots mark descriptions and blue stars the mean of quality estimation for each camera. This makes it possible to see the most distinctive characteristics of cameras' image quality in relation to the other cameras' performance.*

Discussion

The present study shows that even without training the naïve observers can estimate the overall quality of the images consistently. Different observers are also able to describe the pictures so that the descriptions of image quality separate the cameras (see correspondence analysis), which means that the observers use attributes in a similar way for different pictures. The free descriptions can help to complement the MOS results by telling, for example, why the quality of a certain camera is failing.

Here we have shown how the quality descriptions can clarify the subjective measurement data obtained by overall image quality evaluation (MOS). Free descriptions are an easy and fast way to find out, what the naïve observers notice in image quality of certain cameras. This resembles the situation, where typical users are using their imaging devices. With, for example, correspondence analysis it is possible to position the cameras on these dimensions and to see the most characteristic reasons for the image quality score that the camera or the specific test image obtained. This can help to evaluate the overall performance of cameras and to analyze in more detail the complex effects of imaging and camera parameters on subjective image quality.

References

[1] N. Avadhanam, R. V. Algazi, Image compression quality rating using anchored scales, Proc. SPIE. 3308, pg. 44. (1998).

[2] P. G. Engeldrum, Psychometric Scaling: A Toolkit for Imaging Systems Development, (Imcotek Press, Winchester, USA, 2000).

[3] C. Cui, Do Experts and Naïve Observers Judge Printing Quality Differently? Electronic Imaging, SPIE, 5294, pg. 132, (2004).

[4] A. Strauss, & J. Corbin, Basics of Qualitative Research, (2nd ed., Sage Publications, Inc., California, USA, 1998).

Author Biography

Jenni Radun is MA from the University of Helsinki, Finland. She works at the Visual Quality –research group that specializes in the measurement of quality experiences in complex natural images both in printed and electronic imaging environments. Special interest is a deeper understanding of changes experienced in high image quality material, when the higher level psychological dynamics start influencing the interpretations made from the material.