Exploring subjective image quality through isopreference curves

Antti Lehmussola, Pekka Ruusuvuori and Olli Yli-Harja
Institute of Signal Processing
Tampere University of Technology
P.O. Box 553, 33101 Tampere, Finland

Abstract— **Perceptual image quality correlates ineffectively with the traditional error measures. By defining the factors behind perceived image quality, many image processing systems can be optimized. Thus, a great deal of effort has been made for solving this challenging issue. However, considerably less research has been focused on the role of spatial resolution and number of gray levels in subjective image quality. In our study, the relation between these two fundamental image parameters is experimentally defined in terms of perceptual quality. The study was carried out by organizing an experiment where 80 subjects were used. Using the collected data set, the results are illustrated in the form of isopreference curves. Finally, our results are compared with the corresponding ones presented previously. Although the comparison reveals a clear correlation between the results, novel information about the relation between spatial resolution and the number of gray levels is presented.**

I. Introduction

The rapid growth in the number of applications using digital images has emphasized the necessity of studying analytical factors affecting perceptual image quality. By understanding the features characterizing image quality, image processing and compression systems can be highly optimized. However, it is a well-known fact [1] that traditional error measures used in signal processing correlate ineffectively with the perceptual quality of images. Thus, several studies have proposed different metrics for evaluating image quality similarly as human observers do, for example [2], [3], [4].

One of the pioneering studies in the field of subjective image quality assessment was done by Huang who studied the effect of resolution and the number of gray levels in image quality [5]. In the study by Huang, the relation between resolution and the number of gray levels was experimentally quantified in the form of isopreference curves. Although the results are already 40 years old, the publication is still actively referenced, for example by Gonzalez and Woods in [6]. After the study by Huang, less research has been focused on the relation between spatial resolution and the number of gray levels in terms of subjective image quality.

In our paper, we study the relation between spatial resolution and the number of gray levels in terms of subjective image quality, and compare our results to the corresponding ones presented previously in [5]. Especially, we are interested in how the results are affected by a clearly larger group of test subjects and modern computer displays. In Section II we discuss more detailed about the isopreference experiment, and in Section III we describe the results obtained in the test. Finally, concluding remarks and some future directions are given in Section IV.

II. Isopreference experiment

A. Formulation of the problem

The motivation for an isopreference experiment is to study the relation between two (or more) independent variables. Previously, this framework has been utilized in variety of studies, including for example economics and telecommunications. Similarly, the isopreference experiment can be used for studying subjective image quality.

When considering an uncompressed grayscale image, the number of bits needed to store such image is

$$N = L \times L \times B \tag{1}$$

where $L \times L$ is the spatial resolution of the image (for simplification the image is considered as square), and B the number of bits used to represent the intensity of each pixel; that is, the image consists of 2^B graylevels. The basic question then arises: For a given value of N, how the values of L and B should be chosen for obtaining the best possible image quality? Naturally, the decision is a matter of subjective image quality: Whether the spatial resolution has a larger impact on perceptual image quality than the number of gray levels, or on the contrary.

In [5], the relation was determined experimentally using human observers. The experiment was based on three different images containing different amount of details. Test images with varying values for L and B were generated from the original images, and the observers were asked to rank order the images according to the subjective image quality. As a result, so-called isopreference curves were obtained. Isopreference curve is a curve in the $L-B$ plane with each point representing images with equal subjective quality. The experiment presented in [5] was carried out by using approximately 20 subjects, and the images were displayed using photographic prints (Prof. Thomas S. Huang, personal communication, December 7, 2004). The results are illustrated in Figure 1. Following conclusions were drawn from the results. First, the isopreference curves deviate clearly from the constant bit rate. Second, the isopreference curves are highly dependent on the picture characteristics: the curves become more vertical as

0-7803-9134-9/05/$20.00

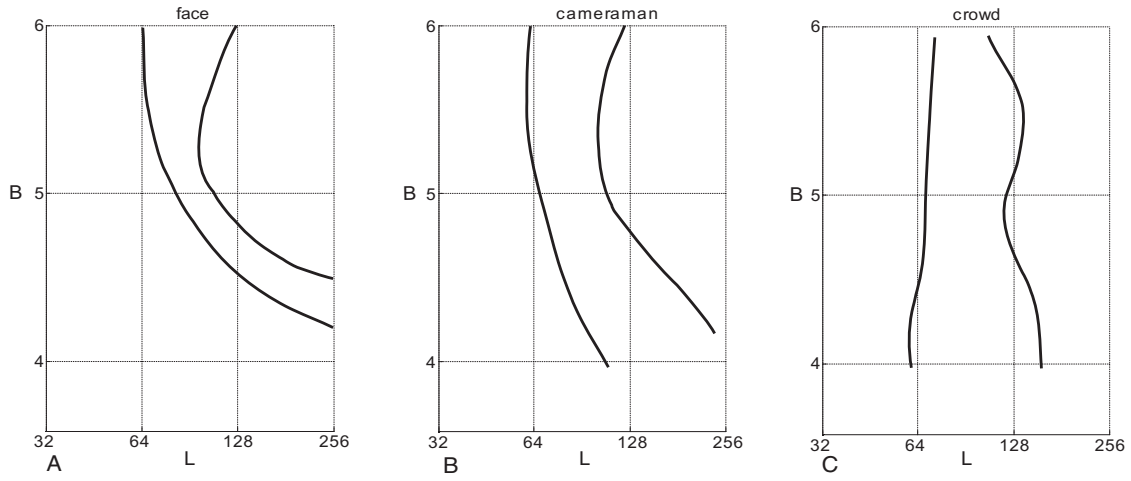

Fig. 1. Restoration of the original results presented in [5] (additional curves discarded). Isopreference curves are curves in the $L-B$ plane with each point representing images with equal subjective quality, where L represents the resolution and B the number of bits for gray levels. The curves are obtained from three test images with different characteristics. First, an image with a human face contained small amount of details (A). Second, an image with a cameraman in the middle contained mediocre amount of details (B). Finally, an image of human crowd contained large amount of details (C).

the number of details increase. The results imply that for the images with minor amount of details, the number of gray levels is more important than for images with large amount of details. For the images with large amount of details, only a few gray levels are needed.

B. Experimental arrangements

In our study, the motivation was to experimentally define how the spatial resolution and the number of gray levels affect subjective image quality. More importantly, we wanted to compare how our results differ from the previous ones. In particular, the usage of computer displays instead of photographic prints was considered as a possible influential factor. We organized an experiment where 80 subjects where used, which was four times larger test group than in the previous study. Since the original test images were not available, we selected our test images similarly as in the isopreference demonstration by [6]. In comparison with the original ones, the modern versions of the test images shared very similar characteristics. More comprehensive description of the images is given in Figure 2. In the previous experiment, several isopreference curves were generated for each image type. However, the more curves are generated, the less data points for each individual curve is obtained if the number of subjects can not be increased. By reducing the number of isopreference curves, the remaining curves will be generated from more data points, and therefore can be considered as more reliable. Thus, we concluded that as the bases for the isopreference curves, two different versions of original images with varying values for L and B will provide sufficient amount of information about each image type. Overall, six different reference images were generated. By selecting the values for L and B from the results in [5], the comparability of our results was guaranteed.

The experiment was carried out using the tool illustrated in Figure 3. The tool provided a graphical user interface for the subjects evaluating the images. Two images were displayed for the observer at a time: a reference image and an adjustable image. The reference image was a random selection from the group of six reference images generated from the original ones. The adjustable image was the same as the reference image with randomly selected number of gray levels. The possible number of gray levels was restricted between 12 and 64 in order to maintain congruity between our results and the previous ones. Furthermore, the resolution of each adjustable image was manually modifiable by the subject. Each subject was asked to select the image resolution so that the perceived image quality corresponds to the quality of the reference image. In addition, a possibility to discard the current image pair was offered for cases where the user could not find a suitable resolution to match the image quality. As a result, a data set consisting of (L,B) point pairs corresponding to images with equal perceptual quality was obtained. Each subject was asked to evaluate altogether 12 different images. Excessive amount of images could have interfered the attentiveness of the subjects. Therefore, the number of images was kept relatively low in order to collect more reliable data points. Finally, as a result 960 data points were collected (including the discarded data points). In order to obtain isopreference curves from the experimental data, we fitted a fourth-order polynomial between the data points.

III. Results

Our current study examines the relation between spatial resolution and the number of gray levels when considering subjective image quality. The results are demonstrated with experimentally quantified isopreference curves for three different image types.

The curves derived from our experimental study are illustrated in Figure 4. Clearly, both isopreference curves in each image category have very similar characteristics. Thus, the

Fig. 2. Since the test images used in the previous experiment were not available, we selected three images having fairly similar characteristics. The images were selected comparably as in the isopreference demonstration in [6]. The selected images represent images with different amount of details, similarly as in the original experiment. First, the image with a human face consists of small amount of details (A). Thereafter, the image with a cameraman consists of mediocre amount of details (B). Finally, the image of human crowd consists of large amount of details (C).

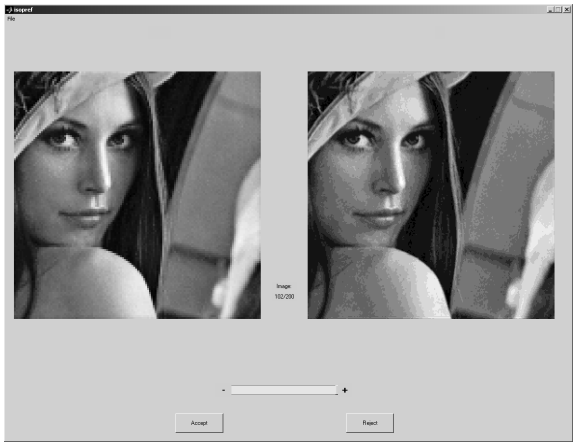

Fig. 3. The graphical user interface of the tool used in the experiment. Two images were displayed for each subject: a reference image (left) and an adjustable image (right). The subjects were asked to change the resolution of the adjustable image so that the two images have equal quality.

curves can be considered to be consisting of enough data points to be reliable. According to our results, the number of gray levels has most importance in case of low information content images, and while the number of details increase, the curves become more vertical. Basically, similar results were also derived in the original experiment [5]. However, our results suggest that the resolution may have a slightly more dominant role as an image quality parameter. Hence, the form of the isopreference curves appears to be less dependent on the image characteristics than the original results suggest.

Interesting differences can be noticed when comparing the endpoints of the isopreference curves. These endpoints correspond to images which have same subjective quality, but largest difference in the number of gray levels. For example, the original results for the low information content image imply that degradation in image quality when decreasing in resolution from 256 to 64 can be neutralized by increasing the number of gray levels approximately from 20 to 64. Interestingly, according to our results the same increase in the number of gray levels is only sufficient for neutralizing the quality loss obtained when decreasing the resolution from 80 to 70. Finally, the original results suggest that the isopreference curves for high information content images should be vertical. When comparing the results in Figure 1(C) and Figure 4(C), it can be seen that our results support this assumption, but with our results the phenomenon is more evident. A possible reason for this is the more excessive data set which can reduce the fluctuations in the results.

IV. Conclusions

The reproduction of the isopreference experiment was carried out in this study. We collected a dataset of altogether almost 1000 samples for six reference images, using 80 different persons for collecting the dataset. By exploring the subjective quality through isopreference experiment, we have provided valuable information about the relation of fundamental image attributes; resolution and number of gray levels, for the perceived quality. Our results resemble those presented in the original study; minor differences are likely to be due to different test images used in the experiment. The large number of subjects is also likely to reduce the fluctuations present in the original curves. However, despite the rather similar results, our results suggest that the resolution could have a more dominating role for the subjective quality of image when compared to that of the number of gray levels. The vital role of resolution should be taken into account when implementing human visual perception driven image compression or transmission methods. However, it has to be emphasized that several factors other than resolution and the number of gray levels also influence the subjective quality of image.

In the future, owing to the emergence of realistic human visual perception based quality metrics, the evaluation of image processing methods based on subjective quality metrics instead of traditional error metrics will increase. The only way to ensure that the quality metrics correspond to the view of human observers is to perform extensive experiments with

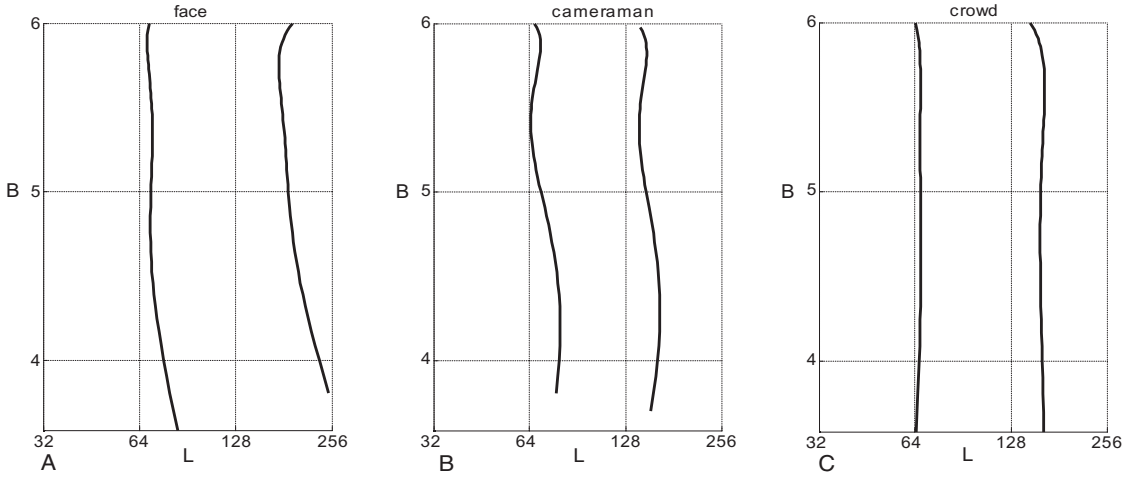

Fig. 4. Isopreference curves obtained from the experiment are constructed by fitting a fourth-order polynomial between the data points. Similarly as in Figure 1, the *L* axis represents the resolution and the *B* axis the number of bits for gray levels. The curves can be considered as reliable, since the both isopreference curves in each image category are very similar. When comparing the derived isopreference curves with the ones presented previously, similar characteristics can be noticed. However, the form of the isopreference curves appears to be less dependent on the image characteristics than the original results suggest.

subjects. As a future direction, we aim at modifying the idea of the original isopreference experiment into a more meaningful and application-oriented experiment, such as studying the perceived quality after filtering, compression or interpolation. In addition, the original experiment could be extended to contain more image parameters than the resolution and the number of gray levels. Even though the role of the number of gray levels seems to play a minor role in image quality, the effect of the number of colors on subjective quality in the case of color images would be an interesting topic for further studies.

V. Acknowledgment

The authors are deeply grateful for Prof. Thomas S. Huang and all the subjects for their help.

References

[1] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1994, pp. 982–986.

[2] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2000.

[3] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.

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

[5] T. S. Huang, "PCM picture transmission," *IEEE Spectrum*, vol. 2, no. 12, pp. 57–63, 1965.

[6] R. C. Gonzalez and R. E. Woods, *Digital Image Processing, Second edition*. Prentice-Hall, New Jersey, 2002.