

THE RELATIONSHIP BETWEEN IMAGE FIDELITY AND IMAGE QUALITY

D. A. Silverstein and J. E. Farrell

Hewlett Packard Laboratories
1501 Page Mill Road
Palo Alto, California 94304 USA
E-mail: amnon@hpl.hp.com

ABSTRACT

Image fidelity (inferred by the ability to discriminate between two images) and image quality (inferred by the preference for one image over another) are often assumed to be directly related. In this study, we investigated the relationship between the perceived image fidelity and image quality of halftone textures. Subjects were asked to rank order a set of printed halftone swatches on the basis of smoothness. They were then asked to reduce the contrast of each pattern until it was at threshold, thus providing an estimate of the pattern's perceptual strength and its discriminability from a non-textured swatch. We found only a moderate correlation between image fidelity and image quality.

1.0 INTRODUCTION

Image fidelity refers to the ability of a process to render an image accurately, without any visible distortion or information loss. If we cannot detect the difference between an original and a compressed image, we conclude that the compression process was visually lossless. It is possible to develop computational measures of image fidelity based on human vision models because these types of judgements depend upon our ability to detect differences between images [1].

There is a natural tendency to confuse image quality with image fidelity. For example, if we cannot detect the difference between a photograph and a digitally printed image, we might conclude that the digital print has photographic image quality. But subjective impressions of image quality are much more difficult to characterize and, consequently, nearly impossible to quantify. It is not difficult to demonstrate that people use multiple visual factors or dimensions in complex non-linear combinations to make judgements about image quality [2]. There are also significant individual differences in their judgements [3]. The two terms, image quality and image fidelity, are often used interchangeably, but they are not the same. Although we often expect fidelity and quality to be closely linked, these two measures are not always positively correlated. An image can sometimes be "enhanced" by a distortion. People may detect the difference between an original and its distorted version and prefer the distorted version over the original.

Typically however, for most noise-like distortions, it is assumed that there is at least a monotonic relationship between the detectability of the noise and the perceived quality of the image. This assumption makes it possible to employ human contrast detection models to make predictions about subjective judgements of image quality. We wanted to determine if one such model, based on a bank of filters sensitive to different spatial frequencies and orientations, could be successfully applied to predict the reduction in image quality due to halftone noise. We estimated the fidelity of halftone patterns by subjects' ability to detect the difference between the pattern and a uniform field with the same space-

averaged reflectance. We estimated the quality of halftone patterns by their rank-ordering of pattern smoothness. We constructed a condition in which image quality judgements could be described by a single dimension and individual differences were minimized. Under these conditions we asked: What is the relationship between image fidelity and image quality?

2.0 IMAGE QUALITY

2.1 Methods

Small printed image swatches (1/2" square) were produced using several different printing technologies and halftoning algorithms. The color of each swatch was selected at random by the computer. In addition to random colors, we included primary and secondary color swatches. This insured that artifacts introduced by the printing mechanism and the halftoning algorithm would have a variety of perceptual strengths. The swatches were pasted to 3" X 5" white paper cards and were broken into 3 decks with approximately 25 cards in each deck. Each card was identified by an arbitrary identification number that was barcoded and printed on the back of the card. 17 subjects were given each of the 3 decks one at a time, and were required to order the decks from the smoothest appearing swatch to the most textured swatch. We attempted to provide a typical viewing environment for printed images. We used uncontrolled illumination that consisted of a combination of overhead fluorescent lights and indirect sunlight. The subjects were not constrained to any viewing distance, and were allowed to manipulate the swatches. They were provided with a work table that was sufficiently large to spread out all of the swatches for simultaneous viewing if they choose to do so. They were given an unlimited amount of time to order the swatches, but most subjects completed each sorting in about 5 minutes.

2.2 Results

We used a one-dimensional scaling procedure to estimate the quality of each swatch in units of standard deviations of preference over the smoothest swatch. We estimated the distance between pairs of swatches as the Z score of the percent of subjects who preferred one of the swatches over the other. We found one-dimensional quality values for each swatch that minimized the RMS error of the distance estimates between all of the swatches. That is, we used the RMS error of the distance estimates as the stress of the system, and we minimized the stress [4].

This procedure was used to test the hypothesis that each sample has an intrinsic one-dimensional quality value, with a single variance for all samples. This value may be based on several different underlying factors, but all subjects may weight these factors in the same way. If different subjects give different weights to several underlying factors or if there are samples with substantially different variances in quality, the stress will be large.

We determined the expected stress by simulation. We simulated 17 subjects making preference judgements with the probability determined by the estimated quality values. We simulated a single quality dimension with equal variance for each sample. We then measured the minimum stress of a fit to the simulated data. We simulated 100 iterations of our experiment and determined the average stress and the standard deviation of the stress. The stress from our experiment was 1.35, and our simulation had a mean stress of 1.33 +/- 0.14. 56 of the 100 simulated runs had lower stress than the actual data, so this was in very good agreement with the hypothesis that the quality judgement of the swatches was one-dimensional.

In each deck there was at least one nearly perfectly smooth sample, so the quality units form an absolute scale with units of standard deviation of preference of the perfect sample over the measured sample.

However, since the data were fit based on relative distances between samples, there is a cumulative error as the sample gets to be many quality units away from the perfect sample. We estimated this cumulative error with the simulation.

To test the precision of the method, we repeated the ranking experiment with a deck composed with several of the cards from each of the previous three decks. Our goal was to see if the quality numbers from the new measurements were the same, even though the comparisons were different. 7 of the 22 samples were estimated as somewhat lower in quality on the second experiment, but this is not unexpected since the errors will be correlated due to the fitting procedure described above. If one sample is estimated as lower quality, all following lower quality samples will be estimated as even lower in quality.

3.0 IMAGE FIDELITY

3.1 Methods

We constructed a viewing device that could optically reduce the contrast of the test swatches from experiment 1. One of the decks from experiment 1 had swatches that had been printed with a continuous tone dye sublimation printer. To introduce visual texture, each swatch was halftoned to fewer levels than the printer was capable of producing. The same printer was utilized to print a second set of swatches that matched the first in color and reflectance, but that were not halftoned and hence had no texture. The optical device mixed the image of the textured swatch with the untextured matching swatch to reduce the pattern contrast.

The device consisted of a corner-cube polarizing beam-splitter. The textured swatch was positioned against one face of the cube, and the nontextured swatch was placed on the perpendicular face of the cube. An incandescent lamp was diffused and illuminated both swatches equally by means of the splitter. The observer viewed the swatches through an aperture positioned so the viewing distance was fixed at an effective 10" with a collimating positive lens. Between the aperture and the cube was a linear polarizing filter. By rotating the polarizer, the observer could control the mixture of the two swatch images, and thus reduce the pattern's contrast. The polarizer had a vernier gradient that was calibrated to percentage mixture of the two images. 5 subjects were instructed to adjust the contrast up and down until the pattern was just barely visible. If the mixture was 10% pattern and 90% blank at the subjective threshold, then we plotted the pattern as 10 times threshold.



Figure 1. A schematic diagram of the contrast reducing device.

3.2 Results

Figure 2 shows the combined results of the two experiments. Each point represents a halftone sample. On the vertical axis, the points are positioned by experiment 1 and on the horizontal axis they are positioned by experiment 2. We averaged the data between subjects after we first normalized each subject's data by a scale factor to account for that subject's criteria. The subject with the lowest criteria was 75% of the average. That is, when this subject adjusted a pattern to be just detectable, she would on average set it to 3/4 the contrast that the other subjects had set. The subject with the highest threshold was 118% of the average threshold.

The error bar was estimated from the data of 5 subjects. A pattern that could not be detected at any contrast would be at 0. However, since the optical device can not measure patterns below threshold contrast, patterns near threshold have arrows on one side of the error bar indicating that their actual visibility may have been lower than 1.

The pattern objectionability was the number of standard-deviations of preference for the blank (non-patterned swatch) over the test swatch (from experiment 1). If the subjects had no preference for the blank over the pattern, the objectionability would be 0. The error bars were obtained by simulating the experiment 100 times.



Figure 2. Pattern objectionability is plotted in units of standard deviations of preference for the non-textured swatch over the measured sample. Error bars were based on the simulation. Pattern visibility is the number of times the pattern was above threshold contrast. Due to the measuring technique, patterns near or below threshold could not be accurately measured (arrowed error bars).

4.0 CONCLUSIONS

From Figure 2 it can be seen that there is at most a weak relationship between fidelity and quality. Surprisingly, the correlation between fidelity and quality is not strong even for some patterns that are very close to threshold. This weak correlation was found despite the use of simple textures and the lack of individual differences. The subjects had good agreement on the quality ranking of the patterns. If the agreement between subjects on quality was worse, as might be expected for more complicated patterns, the relationship between fidelity and quality would become even less direct.

Although these results might seem surprising, they are consistent with other findings that suprathreshold contrast judgements are unrelated to contrast threshold judgements. For example, people will report that two patterns have the same apparent contrast when their physical contrasts are equal, although their contrast thresholds are quite different [5]. Moreover, patients who have documented decrements in contrast sensitivity (amblyopes) are not impaired in suprathreshold contrast judgements [6]. Because it is not possible to predict suprathreshold contrast judgements by contrast detection thresholds, we should not expect human vision models based on contrast detection to predict suprathreshold judgements of image quality.

This does not mean, however, that it is not possible to develop computational measures of image fidelity based on human vision models [1]. Image fidelity judgements depend upon our ability to detect differences between images and this is precisely the type of data one would expect contrast detection models to predict. Contrast detection judgements also play a very important role in image quality assessment. If people detect the difference between an original and a compressed image, the compression is not visually lossless. If people can detect printer banding or halftoning noise in a digital print, they will not confuse it with a photograph. We turn now to determine how well a contrast detection model (based on a bank of filters sensitive to different spatial frequencies and orientations) can predict the halftone contrast thresholds we recorded in our second experiment.

REFERENCES

- 1. S. J. Daly, "The visible difference predictor: an algorithm for the assessment of image fidelity," [SPIE: Human Vision, Visual Processing, and Digital Display III](#), Ed. Rogowitz, Allebach and

[Klein, 1666, pp. 2-15, 1992.](#)

- 2. [A. J Ahumada, Jr.](#) and C. H. Null, "Image quality: A multidimensional problem", SID 92 Digest, pp. 851-854, 1992.
- 3. J. D. Carroll, "Individual differences and multidimensional scaling", in R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*. Vol. 1. Theory. New York,: Academic Press, pp. 105-156, 1972.
- 4. J. B. Kruskal, "Multidimensional scaling by optimizing a goodness of fit to a nonmetric hypothesis", *Psychometrika*, 29, pp. 1-27, 1964.
- 5. M. A. Georgeson, and G. D. Sullivan, "Contrast constancy: Deblurring in human vision by spatial frequency channels", *Journal of Physiology*, pp. 627-656, 1975.
- 6. R. F. Hess, A. A. Bradley and A. L. Piotrowski, "Contrast-coding in amblyopia I. Differences in the neural basis of human amblyopia", *Proceedings of the Royal Society, London*, Vol. 217, pp. 309-330, 1983.