Image Quality and Material Appearance

James A. Ferwerda*
Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

Images are a puzzle for vision scientists because they serve as visual representations of objects while also being objects themselves. We have conducted an experiment that investigates observers' abilities to disregard image distortions to correctly perceive the properties of depicted objects. We studied how well low quality (low contrast, blurry) images conveyed information for surface gloss. We found that gloss perception is not affected as much as might be expected by these distortions. We are using these findings to develop new image quality metrics that distinguish between the signal properties of images and the fidelity of the visual information they convey.

1. Introduction

Realistic images are a puzzle because they serve as visual representations of objects while also being objects themselves. When we look at an image we are able to perceive both the properties of the image and the properties of the objects represented by the image. Research on image quality has typically focused improving the properties of images (resolution, dynamic range, frame rate, etc.) while ignoring the issue of whether the image is serving its role as a visual representation. The danger of focusing exclusively on the properties of the image is that we may miss insights and opportunities about image quality that come from distinguishing between the quality of the imaging medium and the quality of the visual information represented by that medium. Figure 1 illustrates the differences between these two views of image quality.

In this paper we describe an experiment that investigate how well images of different quality convey information about the properties of the objects they represent. The purpose of this work is to understand the relationships between the signal properties of images and the fidelity of the visual information those images convey to human observers. Our goals are to learn more about how images work as visual representations, and to develop more meaningful image quality metrics that better predict how well images with different signal properties serve as visual representations of the objects they depict.

2. Related Work

Measuring image quality is an important aspect of image systems development, and a variety of metrics have been developed for this purpose. *Numerical metrics* such as mean-squared error quantify the distortions in a test image with respect to a real image or a statistical standard. *Perceptual metrics* have been developed that incorporate computational models of human visual processing[1,2]. These metrics often do a better job at predicting perceived image quality than the numerical metrics, however for the most part they still treat images as abstract arrays of pixels.

Fig. 1 Image properties and object properties in imaging: left) High quality grayscale image of a glossy black car on a wet concrete pad. right) Half-toned, printed, and rescanned image. While the quality of the right half-image of is lower than the left, its ability to represent object properties such as the finish of the car and wetness of the concrete is largely the same.

However when we look at realistic images we don't see pixels. Rather, we see objects with recognizable shapes, sizes, and materials, at specific spatial locations, lit by distinct patterns of illumination. Ferwerda and Pellacini[3] introduced the term *functional realism* to describe the idea that radically different renderings (e.g. photographs vs. line drawings) could provide equivalent visual information for given tasks. Ramanarayanan et al.[4] built on this insight and developed a new measure of image quality called *visual equivalence*.

The common thread in this work is an understanding of the value of distinguishing between images as signals that reproduce patterns of light and images as messages that convey visual information to observers. This distinction provides a new perspective on image quality that we investigate in the following experiment.

*mailing: Lomb Memorial Drive, Rochester, NY 14623, USA
email:jaf@cis.rit.edu

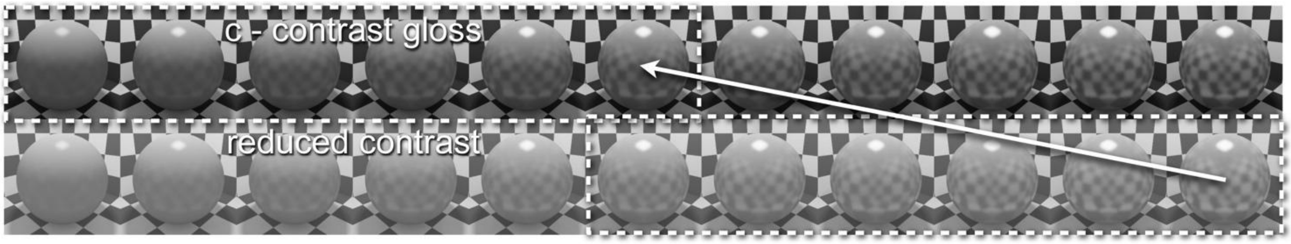

Fig. 2 Images used in the experiments, “c” set. The contrast gloss of the balls ranges from 0.019 (low) to 0.190 (high). The top row shows the normal contrast images. The bottom row shows the low contrast images. The reflections in the six glossiest balls in the lower row have approximately the same image contrast as the reflections in the six least glossy balls in the top row.

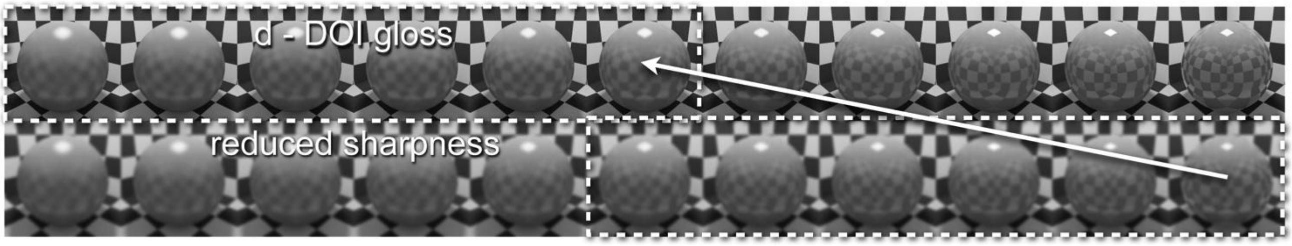

Fig. 3 Images used in the experiments, “d” set. The distinctness-of-image gloss of the balls ranges from 0.9 (low) to 1.0 (high). The top row shows the normal sharpness images. The bottom row shows the low sharpness images. The reflections in the six glossiest balls in the lower row have approximately the same edge profiles the reflections in the six least glossy balls in the top row.

3. Experiment

We designed an experiment to investigate how well images of different quality convey information about the properties of the objects they represent. In particular we studied how well observers were able to perceive the gloss of objects represented by normal and low quality (low contrast, blurry) images.

3.1 Stimuli

The stimuli used in the experiment are shown in Figures 2 and 3. They are computer graphics renderings of a ball in a checkerboard box with an overhead light source. The materials in the scene were described using the Ward light reflection model[5]. In previous work Pellacini et al.[6] developed a gloss model that showed that the perceived gloss of objects rendered in images was well described by two perceptually uniform parameters, c – contrast gloss, and d – distinctness-of-image gloss, that could be directly related to the parameters of the Ward model.

This model was used to describe the gloss properties of the balls in the scene. In the “c” set (Figure 2 top row) the contrast gloss varied in equal steps from 0.019 on the low end to 0.190 on the high end, “d” was fixed at 0.93. In the “d” set (Figure 3 top row) the distinctness-of-image gloss varied in equal steps from 0.9 to 1.0. “c” was fixed at 0.087. The scene was rendered at 300x300 pixels using the Radiance rendering system[5].

Two manipulations (contrast reduction, low pass filtering) were used to produce the “low quality” image sets. To create The “low contrast” image set had its contrast reduced by raising the images’ black level in Photoshop so that the checkerboard reflection in the glossiest ball in the low contrast set had the same Michaelson contrast as the checkerboard reflection the middle gloss ball in the normal contrast set. Similarly, the “low sharpness” image set (Figure 3 bottom row) was created by applying a Gaussian blur filter so that the edge profile of the checkerboard reflection in the glossiest ball in the low sharpness set was the same as the profile in the middle gloss ball in the normal sharpness set.

3.2 Methods

Using these image sets we performed a gloss scaling experiment to investigate the effects of image contrast and sharpness on gloss perception. The interface used in the experiment is shown in Figure 4. The left and right halves of the Figure show sample trials from the “c” and “d” conditions that were run as separate studies. In each study, each trial was composed of a pair of screens shown to the observer in sequence. At the start of a trial a reference screen appeared which showed one of the balls from the either the normal or low quality image sets in the upper half of the screen. Only the six most glossy balls in each set were used as reference balls for reasons that will be explained below. Observers were instructed to “Look at the reference ball and observe how glossy it is.” then “Click (here) when you are ready to continue.” On clicking, the reference screen disappeared and was replaced with a test screen that showed three balls selected

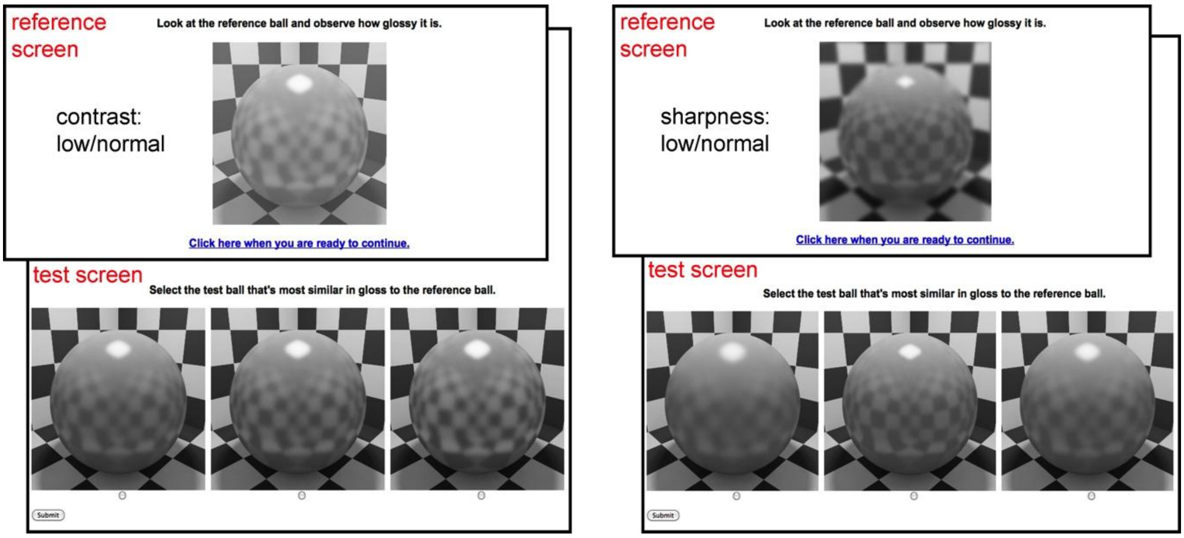

Fig. 4 Sample trial screens from the experiment: The left panel shows a trial from the "c" condition. The right panel shows a trial from the "d" condition. At the start of a trial the reference screen appeared observers viewed the object/image and then clicked the link. The reference screen then disappeared and the test screen appeared. The observers' task was to select the test ball with the same gloss as the reference ball. Both normal image vs. normal image and low quality vs. normal image trials were presented.

from either the normal or low quality image sets. The gloss values of the test balls were: 1) the same as the reference ball; 2) two or three steps lower than the reference ball (presentations were split across observers and averaged); and 3) six steps lower than the reference ball. The left/center/right positions of the test ball images were randomized. The observers were instructed to "Select the test ball that's most similar in gloss to the reference ball." by clicking on one of the radio buttons beneath each image. The observers then clicked a "Submit" button and the next trial sequence began.

The reference and test ball gloss values were selected with the following rationale. The normal reference vs. normal test trials were created to establish a baseline for performance in the sequential matching task. The expectation was that observers would most often select the test ball with the same gloss as the reference ball, but that there would be some variance, because in most cases the three test balls were just barely noticeably different in gloss. The low quality vs. normal trials were created to test the hypothesis that observers would be more accurate in judging the gloss properties of the balls if the images were presented sequentially rather than side-by-side. In these trials the test images present the observer with three choices. In one case they could choose a ball that has the same gloss properties as the reference ball (material match), in another case they could choose a ball whose image has the same contrast or sharpness as the reference ball (image match), and in the final case they could choose a ball/image with properties midway between the material and image matches.

Forty observers participated in the experiment (twenty in each condition), which was conducted online using the Amazon Mechanical Turk system[7]. While running experiments using the Turk system offers no control over viewing conditions or observer characteristics, if the observers perform consistently it suggests that the findings will be robust in real-world applications.

3.3 Results and Discussion

For each of the six reference balls tested, the frequencies with which the observers selected the material match, the image match or the mid-value match were tallied and averaged to estimate the perceived gloss of each reference ball in the normal and low image quality conditions. The results are summarized in Figures 5 and 6.

Figure 5 shows the results for the "c" condition. The upper line indicates the matching value that would be obtained if observers made an ideal material match between the reference and test balls, the lower line indicates the matching value that would be obtained if observers made a perfect image contrast match. The upper curve shows the result for the normal reference vs. normal test trials. In this baseline condition, observers strongly tend toward making material matches. The more interesting result is shown by the lower curve. This curve shows the average matching values obtained in the low contrast vs. normal trials. Here while the average match values are lower than in the normal vs. normal condition, in most cases the differences are not significant, and more importantly, the matching values are much closer to the material match line than the image contrast match line. As proposed earlier, this finding suggests that observers'

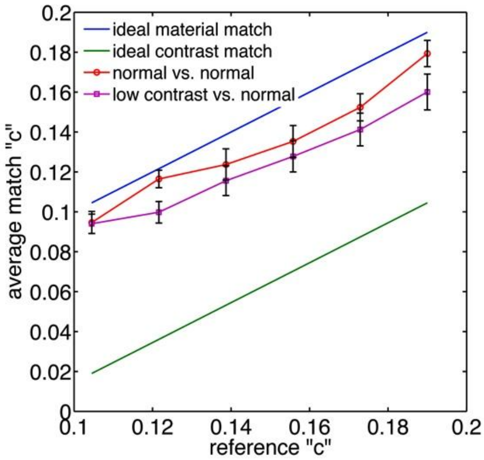

Fig. 5 Results, "c" condition. Matching "c" values for objects with different reference "c" values are plotted as functions of image contrast. Note that both the normal image vs. normal image trials and low contrast image vs. normal image trials tend toward producing material matches rather than image contrast matches.

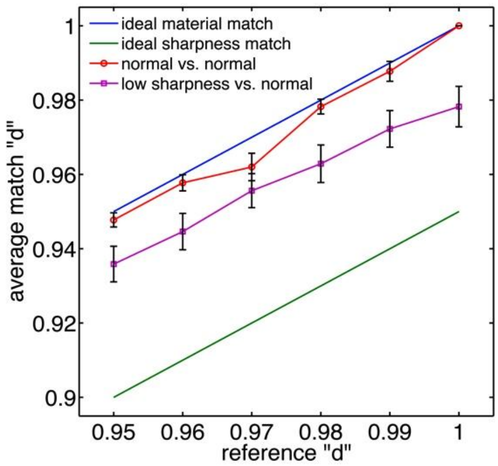

Fig. 6 Results, "d" condition. Matching "d" values for objects with different reference "d" values are plotted as functions of image sharpness. Note that both the normal image vs. normal image trials and low sharpness image vs. normal image trials tend toward producing material matches rather than image sharpness matches.

judgments tend toward the veridical (matching object gloss rather than image contrast).

Figure 6 shows the results for the "d" condition. Similar to the results for the "c" condition, the average matches for the normal vs. normal trials are close to material matches and the matching values for the low sharpness vs. normal conditions while significantly below ideal material matches are still closer to this standard than to the image sharpness match standard. Similar to above this suggests that while observers are influenced by the properties of the images, they tend to choose test images that match in terms of object features (gloss) rather than image features (sharpness).

4. Conclusion

In this paper we described an experiment that explored how well images of varying quality serve as visual representations of the objects they depict. Under the conditions tested, we found that observers behave as if they are able to some degree to "see through" the contrast and sharpness limitations of images to perceive the object properties the images represent. This work is very preliminary, and there is much more that can be done, but the findings suggest that in understanding image quality that there are useful distinctions to be made between the quality of the imaging medium and the quality of the visual information represented by that medium.

Acknowledgements

This work was supported by the National Science Foundation through award IIS-0113310 to the author.

References

1) Daly, S. (1993) The visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, 179–206.
2) Wang, Z., Bovik, A. C., Sheikh H. R., and Simoncelli, E. P. (2004) Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612.
3) Ferwerda, J.A and Pellacini, F. (2003) Functional difference predictors (FDPs): measuring meaningful image differences. Asilomar Conference on Signals, Systems, and Computers, 1388-1392.
4) Ramanarayanan, G., Ferwerda, J.A., Walter, B.J. and Bala, K. (2007) Visual Equivalence: towards a new standard for image fidelity. ACM Transactions on Graphics, 26(3), (SIGGRAPH '07), 1-11.
5) Ward, G.J. (1994) The RADIANCE Lighting Simulation and Rendering System. Proceedings SIGGRAPH '94, 459-472.
6) Pellacini, F., Ferwerda, J.A. and Greenberg, D.P. (2000) Toward a psychophysically-based light reflection model for image synthesis. Proceedings SIGGRAPH '00, 55-64.
7) Amazon Mechanical Turk (MTurk) https://www.mturk.com