The medium and the message: a revisionist view of image quality

James A. Ferwerda
Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, NY, USA*

ABSTRACT

In his book “Understanding Media” social theorist Marshall McLuhan declared: “The medium is the message.” The thesis of this paper is that with respect to image quality, imaging system developers have taken McLuhan’s dictum too much to heart. Efforts focus on improving the technical specifications of the media (e.g. dynamic range, color gamut, resolution, temporal response) with little regard for the visual messages the media will be used to communicate. We present a series of psychophysical studies that investigate the visual system’s ability to “see through” the limitations of imaging media to perceive the messages (object and scene properties) the images represent. The purpose of these studies is to understand the relationships between the signal characteristics of an image and the fidelity of the visual information the image conveys. The results of these studies provide a new perspective on image quality that shows that images that may be very different in “quality”, can be visually equivalent as realistic representations of objects and scenes.

Keywords: McLuhan, image quality, image fidelity, visual perception, visible difference predictors, visual equivalence

1. INTRODUCTION

In his influential 1964 book “Understanding Media”[1], social theorist Marshall McLuhan declared: “The medium is the message.” The idea that McLuhan put forth is that print-based and electronic media technologies have different capabilities and limitations, and this influences the messages that can be communicated. McLuhan believed that the dominant media used by a society shape public discourse, and this in turn determines the possible paths for societal development. McLuhan’s key point is that the medium by which a message is communicated can be as important as the message itself.

The thesis of this paper is that with respect to image quality, imaging system developers have taken McLuhan’s dictum too much to heart. Efforts focus on improving the technical capabilities of the media (e.g. resolution, frame rate, dynamic range, color gamut), with little regard for the messages (image content) that the media will be used to communicate. The faith is that if the image signal is ideal then the message will be conveyed with high fidelity. This approach seems logical and has strong mathematical support from the fields of signal processing and information theory. However the danger of focusing exclusively on the signal properties of images is that we may miss insights and opportunities that come from distinguishing between images themselves and the visual information the images represent.

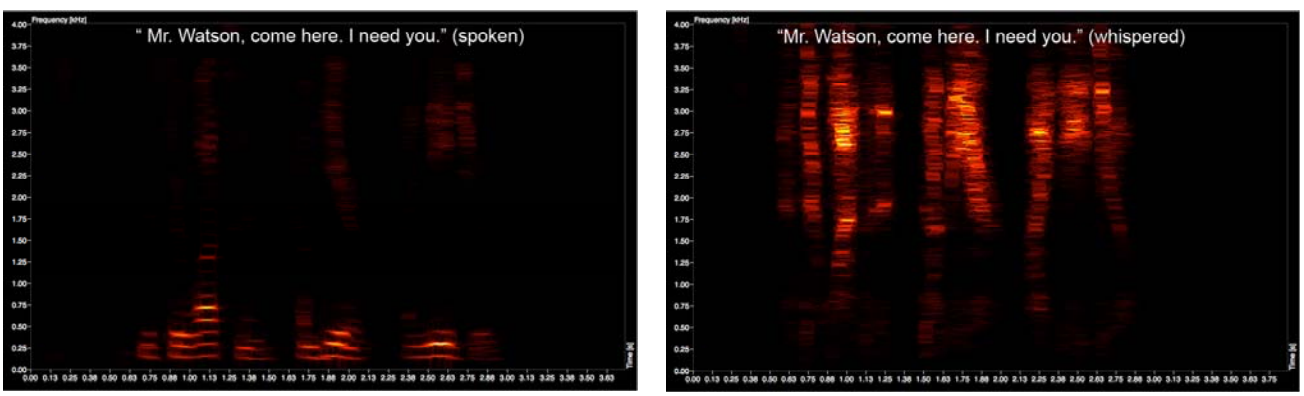

Figure 1: Spectrograms of the phrase “Mr. Watson, come here. I need you.” Spoken (left). Whispered (right). Note the differences in the range and structure of the audio frequency spectra. The audio signals are distinctly different but the message carried by the signals is largely the same.

* jaf@cis.rit.edu, http://www.cis.rit.edu/jaf

Figure 1 illustrates the importance of distinguishing between the signal and the message in auditory communication. The left and right panels show audio spectrograms[2] of the phrase: "Mr. Watson, come here. I need you.", famously uttered into the first telephone, by Alexander Graham Bell. The phrase on the left was spoken in a normal voice. The phrase on the right was whispered. Note the differences in the energy distributions of the two spectrograms. The spoken record consists mainly of audio frequencies below 1 Khz and the audio spectra show distinct harmonic structures. The whispered record on the other hand has very little energy below 1Khz, and consists of unstructured broadband audio spectra. Here is a case where the audio signals are dramatically different, yet the message communicated is effectively the same. Certainly there are differences in the qualities of the signal (forcefully spoken vs. whispered) that may say something about the condition or intention of the speaker, but these differences do not fundamentally affect the linguistic content of the message.

Figure 2. Signals and messages in imaging: a) High quality grayscale image of a glossy black car on a wet concrete pad. b) Halftoned, printed, and rescanned version of the image on the left. Note that while the visual "quality" of the image on the right is lower, its ability to represent important scene properties such as shape and material is largely the same as the image on the left.

Figures 2 illustrates that similar distinctions can be drawn between the signal properties of the medium and the fidelity of the message in imaging. Figure 2a is a grayscale photograph of a sports car parked on a concrete pad[3]. Both the car and the pad show distinct reflections of the surrounding environment that indicate that the car is shiny and the concrete pad is wet The levels and contrasts of the patterns of light in the image also indicate that the car is black (or a dark color) and the concrete pad is a lighter shade. Figure 2b shows the same subject represented by a halftoned image that was printed and rescanned (to approximate a newspaper print). While this image is clearly different than the one in 2a, and in conventional terms one would say that the quality of this image is lower, as a visual representation of the scene this image is quite comparable to 2a in that we can still perceive important properties of the objects such as the shape and shininess of the black car and the wetness and lightness of the concrete pad.

We are currently conducting a program of research[4-6] to investigate the visual system's ability to "see through" image distortions such as those in Figure 2b to perceive the visual messages (object and scene properties) the image represents. The goal of this work is to understand the relationships between the characteristics of image signals and the fidelity of the visual information the images convey. Our goal is to broaden the concept of image quality to identify classes of visual representations that may look distinctly different from one another, but are visually equivalent as representations of objects and scenes.

2. VISUAL EQUIVALENCE

Measuring image differences is an important aspect of image quality testing, and a variety of metrics have been developed for this purpose. *Numerical metrics* measure physical differences between a reference image and test image. Well known numerical metrics include mean squared error (MSE) and peak signal to noise ratio (PSNR). Although these metrics are easy to compute, they often do not correlate well with observers' judgments of image differences. For this reason, *perceptual metrics* incorporate computational models of human visual processing. Typically in these metrics, visual models are used to represent an observer's responses to a reference and test image and then these responses are compared to identify visible differences. Popular perceptual metrics include Daly's Visible Differences Predictor

(VDP)[7], the Lubin/Sarnoff model[8], and the Structural Similarity Metric (SSIM)[9]. These metrics typically do a better job at predicting perceived image quality. However current perceptual metrics have an interesting limitation that is illustrated in Figure 3.

Figure 3a and 3b show two computer-generated images of a tabletop scene. Figure 3a was rendered using path tracing, a physically accurate but computationally intensive graphics algorithm. Figure 3b was rendered using environment mapping, a fast but approximate rendering algorithm that uses an image of the surround rather the surround itself to illuminate the objects on the tabletop. One consequence of environment mapping is that illumination features such as surface reflections are warped from their true configuration. This can be seen by comparing the images reflected by the two teapots.

a) Path traced image b) Environment mapped image c) Visible differences map

Figure 3. a,b) Computer graphics images rendered with different reflection algorithms and c) the output of a VDP metric showing areas of visible difference. Note that while the images are visibly different, they are similar in quality, and convey equivalent information about object appearance.

If we take the path traced image as the reference, and the environment mapped image as the test, and process the images with Daly's VDP, it produces the difference map shown in Figure 3c which correctly indicates that the images *are* visibly different (green and red pixels 75% and 95% probability of detection respectively). However a key question is: *Are these meaningful image differences?*

When we look at 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. From this perspective the two images shown in Figure 3 are much more similar than they are different. For example, the shapes, sizes, materials and locations of the objects appear the same in both images, and the scene lighting looks the same. Although the images are *visibly different* they are *visually equivalent* as representations of object appearance. The existence of images like these has prompted us to develop a new kind of image difference/quality metric that can predict when different classes of image transformations produce images that are visually equivalent[4-6].

The concept of visual equivalence is based on the premise that two visibly different images can convey the same information about object and scene properties to a human observer. To develop metrics for predicting when images will be visually equivalent, we need to understand the factors that determine object appearance.

The appearance of an object is based on the images it reflects to our eyes. For a static object and fixed viewpoint, these images are determined by the object's geometry and material, and the way it is illuminated. To begin to quantify the phenomenon of image equivalence we decided to study equivalence across two kinds of illumination transformations (blurring and warping) for objects with varying geometries and material properties.

2.1 Stimuli

We first created a set of computer graphics stimulus images that would allow us to systematically explore the visual interactions between object geometry, material, and illumination. Figure 4a shows representative images from our stimulus set which showed a bumpy ball-like test object on a brick patio flanked by two pairs of children's blocks. The four object geometries (G0-G3) were judged in pre-testing to be equally spaced with respect to surface "bumpiness". The four materials (M0-M3) represented rolled aluminum with different degrees of microscale roughness.

Since recent studies have demonstrated the importance of real-world illumination for the accurate perception of shape and material properties we lit our model using Debevec's "Grove" HDR environment map[10] that captures the illumination field in the Eucalyptus grove at UC Berkeley. We chose this map in particular, because Fleming et al.[11] found that it allowed subjects to most accurately discriminate material properties. Starting with the original "Grove" map as the reference then generated two sets of transformed maps (blurred, warped). The top two rows in Figure 4b show the five levels of the blurred map set and its effects on the appearance of the G1/M0 object. The third and fourth rows show the warped map set and its effects on the same object.

Images were rendered at 484x342 pixels as high dynamic range (HDR) floating point images using a custom-built physically-based Monte Carlo path tracer that incorporated the Ward[12] light reflection model. The HDR images were tone mapped using a global sigmoid tuned to the characteristics of the display. Each image subtended approximately 12 degrees of visual angle.

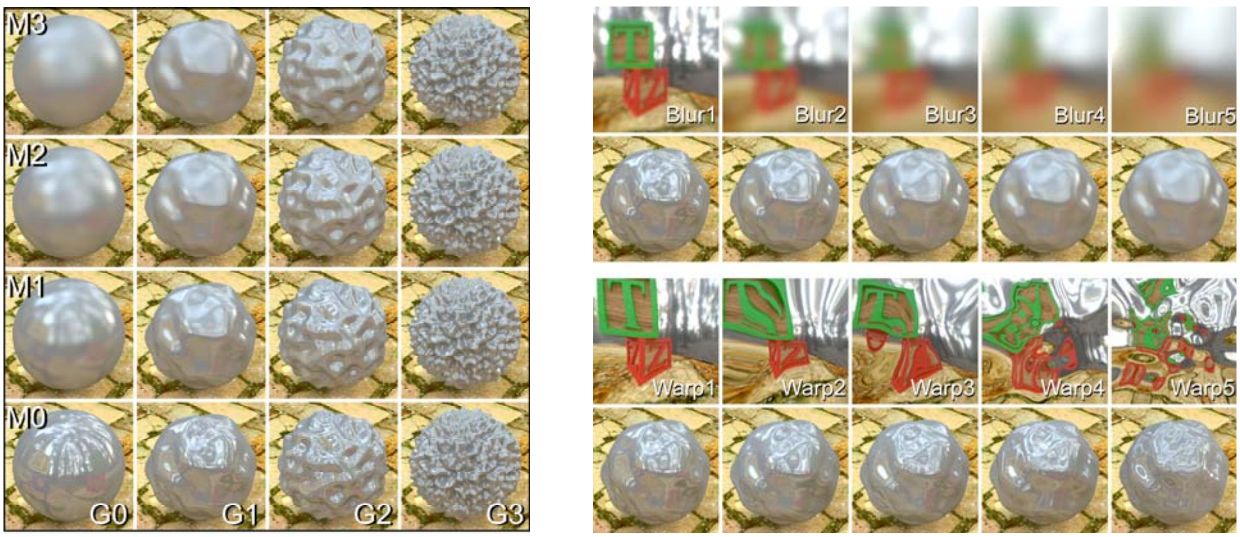

Figure 4. a) The geometries and materials of the objects used in the experiments. Parameters were chosen to be perceptually uniform in both surface “bumpiness” and surface reflectance; b) The two classes of illumination transformations used in the experiments (blur and warp). The upper and lower panels show direct views of the blurred and warped illumination maps and their effects on the appearance of a representative object (G1/M0).

2.2 Procedure

The images in the stimulus set were presented to subjects in pairs. In some conditions a third reference image was shown above the test pair. *In all cases the test pairs showed objects with identical shapes and material properties* (the G/M combinations shown in Figure 4a). In each case one of the images was rendered using the reference illumination map, and the other was rendered using one of the transformed maps (Blur1-5 or Warp1-5 as shown in Figure 4b). An experiment consisted of four related questions/tasks asked about the image pairs.

> *“Which test image is the same as the reference image?”* The intent of this task was to determine when images rendered with the transformed maps were visibly different (in the VDP sense) from images rendered with the reference map.
>
> *“Are the left and right test objects the same shape?”* The intent of this task was to determine if the transformed maps produced illusory changes in the apparent shapes of the objects.
>
> *“Are the left and right test objects made of the same material?”* The intent of this task was to determine if the transformed maps produced illusory changes in the apparent material properties of the objects.
>
> *Which test object is lit the same as the reference object?”* The intent of this task was to determine if subjects could use surface reflection patterns to detect differences in scene illumination.

Overall 15 subjects (ages 20 to 50) participated in the experiments. Some had technical backgrounds, but none in imaging. All were naive to the design and purpose of the experiments and had normal vision.

Note that the four tasks can be divided into two conceptual categories. In the image difference task subjects are being asked to report on detectable *image* differences. In the shape, material, and illumination difference tasks subjects are being asked to report on detectable *object* differences. We chose these tasks because they should allow us to dissociate the effects of image differences on image and object appearance and quantify when different configurations of object geometry, material and illumination produce images that are visually equivalent.

2.3 Results

The results of the experiments are summarized in Figure 5. Each panel shows the set of objects tested (G0-G3, M0-M3), and the upper and lower strips show how observer's judgments changed for different levels of illumination map blurring (Blur1-5) and warping (Warp1-5). The results fall into three categories. In general green symbols are good and red symbols are bad.

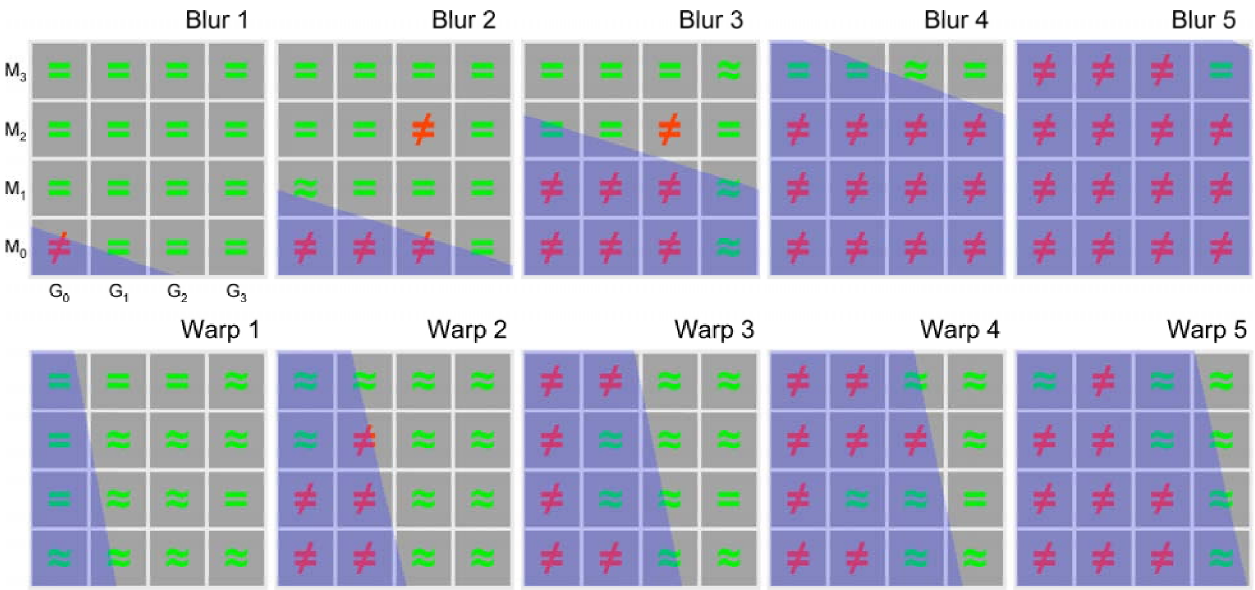

Figure 5. Results of the experiments: Each panel represents the objects tested (G0-G3, M0-M3) (see Figure 4). The upper and lower strips show the results for different levels of the blur and warp illumination transformations. Overall the results fall into three categories: equality, non-equality, and equivalence. The borders between the blue and gray regions represent SVM cuts through the dataset that separate the good cases (green: equal, equivalent) from the bad cases (red: not-equal). This model seves as the basis of a visual equivalence predictor (VEP).

> *Equal:* When subjects reported that the reference and test images were indistinguishable we labeled the images as equal (green = signs). Note that for low levels of blur the image differences were often undetectable and the images appeared identical.
>
> *Not-Equal:* On the other hand when subjects reported that the reference and test images were visibly different, and also reported that the objects also looked different we labeled the images as not equal (red ≠ signs). Note that the number of non-equal cases increases monotonically with the magnitude of the blur and warp transformations.
>
> *Equivalent:* Finally, when subjects reported that the reference and test images were visibly different but also said that the objects represented by the images appeared the same (same geometry and material, no clear differences in illumination) we labeled the images as equivalent (green ≈ signs). Note that while there are few equivalent cases for the blur transformation, there are cases of equivalence at all levels of the warp transformation, even the most severe.

What these results show is that there is a significant class of conditions (indicated by green symbols) where the images rendered with the transformed illumination maps are either equal or equivalent to the reference renderings as representations of object appearance. While existing visible difference metrics (VDPs) could predict the cases of equality, they would not identify the much larger set of visually equivalent images.

2.4 A Visual Equivalence Predictor

To take advantage of this newfound latitude in perceptually acceptable image distortions we developed a new kind of image metric: the *visual equivalence predictor* (VEP). To achieve this we first used Support Vector Machines to classify the experimental results into "good" and "bad" categories. The classification planes are illustrated by the blue-shaded regions in Figure 5. This process yields a predictor that can determine visual equivalence for the images in our test set. However to be useful, we need be able to predict equivalence for images of novel objects and scenes. For novel geometries we characterize the average surface normal variation for the objects in our test set (G0-G3) and map the normal variations of new objects into this space. For novel materials, we fit surface reflectance data with the Ward light reflection model and cast the parameters into our M0-M3 material space. Finally for illumination, our only requirement is that the illumination fields have approximately "natural" image statistics ($\sim 1/f^2$ power spectrum).

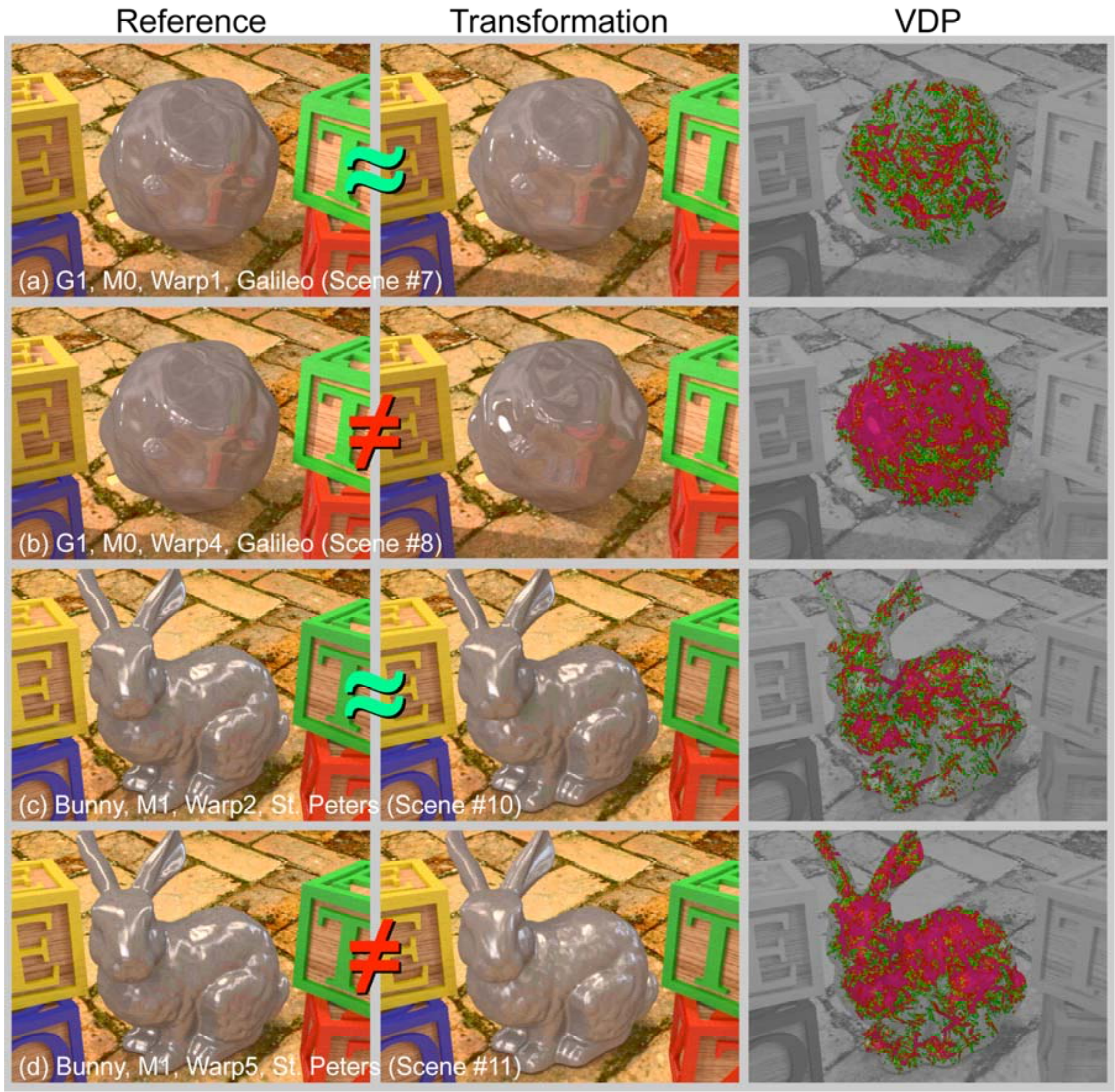

Figure 6. Selected images from the validation experiment. The reference images (left), test images (middle), visible difference (VDP) maps (right). Symbols on each image pair indicate whether they were seen/predicted to be visually equivalent despite being visibly different.

To test the predictive power of the VEP, we ran a confirmatory experiment where we created reference and test images of 14 novel scenes, ran them through the predictor and also had subjects judge them using the same procedure employed in the main experiment. Ten new subjects participated. The VEP correctly predicted the result in 13 out of 14 cases (being overly conservative in one case), and was able to predict both equivalence and non-equality. Selected results are shown in Figure 6.

The result of the efforts described above is a new foundation for image difference/quality metrics: visual equivalence. Images are visually equivalent if they convey the same information about object appearance even if they are visibly different. We believe that visual equivalence is a novel approach to quantifying image quality that goes significantly beyond existing metrics by taking advantage of the limits of visual coding of object appearance and leveraging the fact that some classes of image differences do not matter to human observers.

3. WORK IN PROGRESS

We are working to further explore and understand the phenomenon of visual equivalence. Work in progress includes an investigation of the role of natural illumination statistics in visual equivalence and studies of visual equivalence in dynamic scenes.

3.1 Visual Equivalence and Illumination Statistics

In the experiments described in the previous section, two transformations (blurring and warping) were applied to the illumination maps used to light the objects. A close look at the experimental results (Figure 5) reveals that the blur transformation led to many more "not-equal" cases (39 out of 80) than the warp transformation (27 out of 80). Why do some illumination map transformations lead to more or less visual equivalence than others? One way to gain insight into

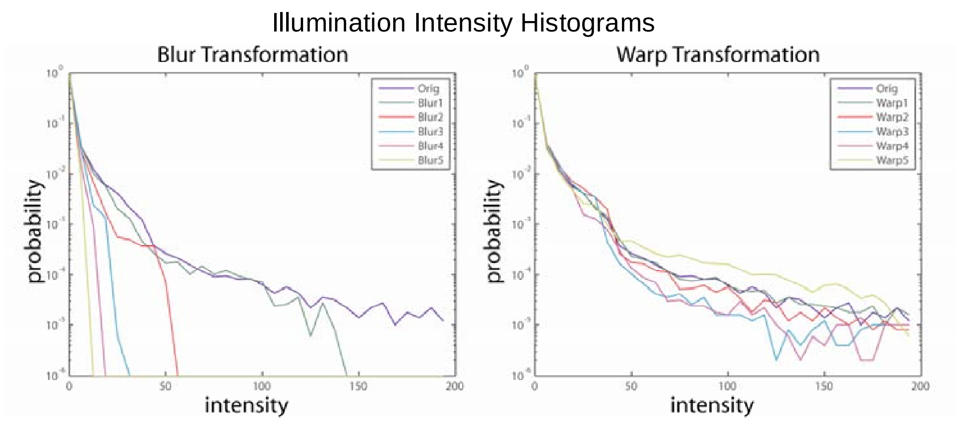

Figure 7. Illumination intensity histograms for the blur and warp transformations. Note that blurring significantly reduces the original intensity dynamic range, while warping leaves it relatively intact.

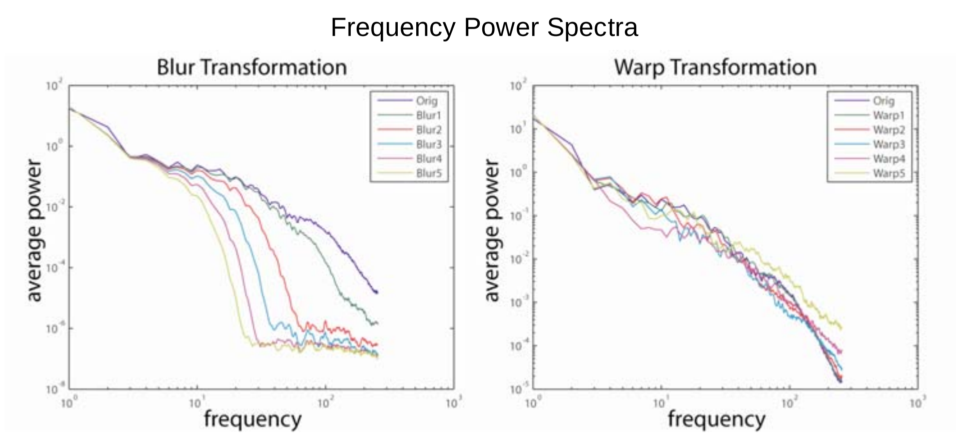

Figure 8. Frequency power spectra for the blur and warp transformations. Note that blurring effectively low pass filters the original spectrum while warping leaves it relatively constant.

this question is to look at natural illumination statistics. Recent work by Dror et al.[13] has shown that natural illumination maps, such as the one we used, exhibit many statistical regularities. One question is how these statistics are affected by the transformations we applied? In Figures 7 and 8, two standard statistical measures from Dror's paper aere plotted: illumination intensity and frequency power spectra for the blurred and warped illumination maps used in the experiments. Notice first that the level of blurring increases, both the number of high intensity values in the illumination map (Figure 7a) and the average power at higher frequencies (Figure 8a) decrease; this is to be expected because blurring effectively low pass filters the illumination map. The warp transformation however, has a different effect. Both the intensity and frequency power spectra (Figures 7b, 8b respectively) stay relatively constant regardless of the magnitude of the warp. This invariance may help to explain why the warp transformation yields greater equivalence than the blur transformation, however further studies are necessary. Investigations along these lines[14,15] may allow us to predict other computationally useful classes of image transformations that produce images that are visually equivalent to reference renderings.

3.2 Visual Equivalence in Dynamic Scenes

One of the potential applications for visual equivalence metrics is the development of perceptually-based acceleration schemes for computer graphics rendering. The equivalence metric presented in the previous section applies to static images. While this has some utility, it would be a real advantage if the equivalence concept applied to dynamic scenes as well so it could be used to accelerate animation renderings. At first glance it is not clear whether object or camera motions should enhance or destroy equivalence relationships between reference and transformed images.

Figure 9. Reference sphere showing interior (left) cathedral and exterior (right) courtyard scenes used to illuminate the test objects.

We have recently studied visual equivalence in dynamic scenes[6]. An experimental paradigm similar to the one described above and in [4] was used, however the stimuli were animations rather than static images. Nine objects (all combinations of G1-G3 and M0-M2 from [4]) and three classes of motion were studied (camera rotation, object rotation, and object translation). The scene consisted of the interior and exterior courtyard of a cathedral (shown in Figure 9).

Equivalence under rotation: In the object and camera rotation conditions the objects were inside the cathedral and were lit with illumination maps created by rendering a spherical HDR image of the cathedral model. The reference illumination map was then warped using warp levels 1-4 from [4]. Subjects were presented with triplets or pairs of animations (+/- 15^o rotations, 30^o/s velocity, unmatched viewpoints) and in separate studies were asked four questions about each set: 1) Indistinguishability: Which of the test movies is the same as the reference movie?; 2) Shape difference: Do the objects (in the movies) have the same shapes?; 3) Motion difference: Are the objects moving in the same way?; and 4) Illumination difference: Which of the test objects is illuminated the same as the reference object?

The results of the rotation experiments are summarized in Figure 10a,b. Overall the results are consistent with those found in [4]. Under the conditions tested the reference and warped renderings are mostly equal or equivalent indicating that rotation of either the object or the camera does not destroy equivalence at the warp levels used.

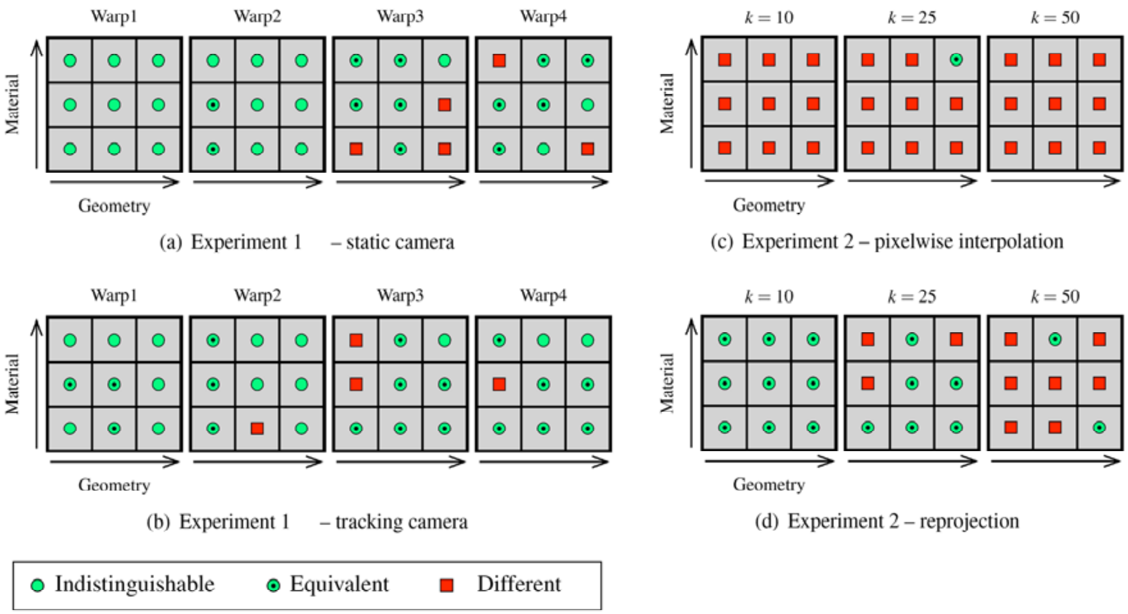

Figure 10. Results of dynamic visual equivalence experiments. Experiment 1 – rotation. Experiment 2 – translation.

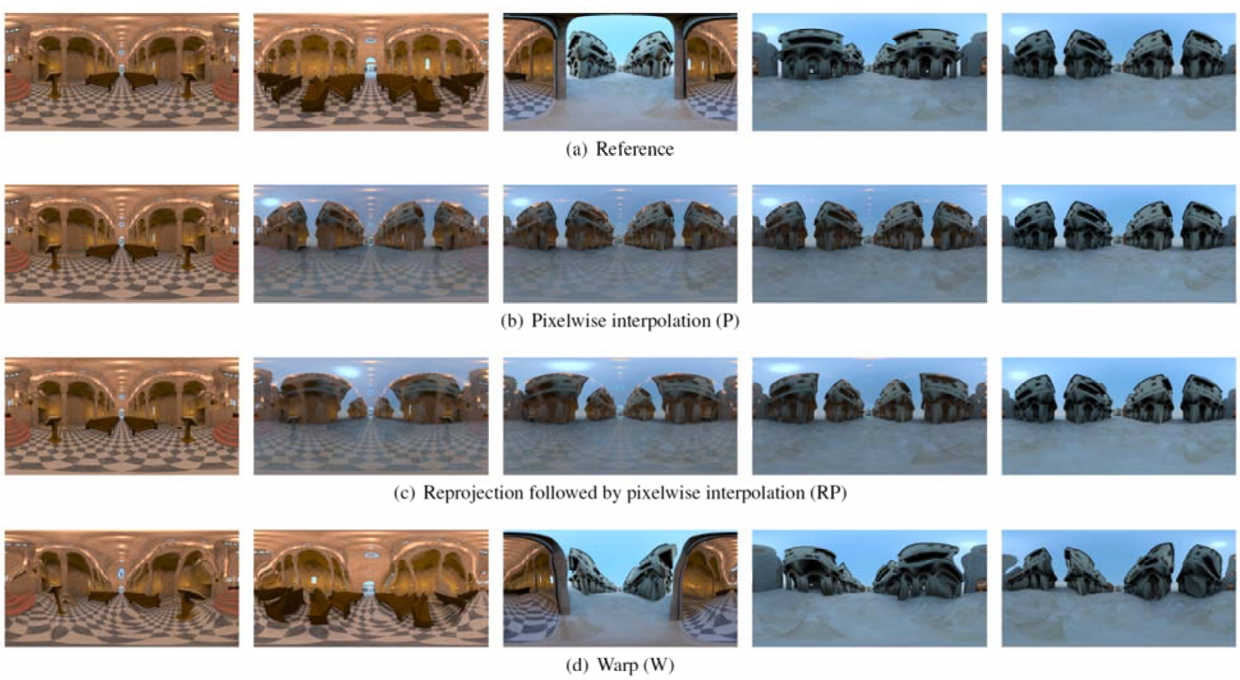

Figure 11. Example illumination transformations. The images show latitude-longitude projections of the illumination maps used to light the test objects. Each row illustrates one kind of illumination transformation used in the translation experiment. The columns show five keyframes along the entire translation path. In the actual animations illumination map keyframes were generated at much finer intervals.

Equivalence under translation: In the translation conditions the object flew without rotation from inside the cathedral to the outside courtyard. Distinct illumination maps were created for rendering each frame of the animations using one of the four schemes illustrated in Figure 11. In the reference condition (11a) a new map was generated for each frame. In the pixelwise interpolation condition (11b, P), keyframe maps were generated at regular intervals and the pixel values were linearly interpolated to create the intermediate maps. In the reprojection condition (11c, RP) values from the keyframe maps were projected onto a sphere located at the intermediate location and this map was then used to illuminate the object. In the warp condition (11d, W) warp transformation level 2 from [4] was applied to each reference map with an added temporal coherence constraint. Maps were generated and animations were rendered at three different keyframe intervals (k= 10,25,50 frames). Because the results of the rotation study showed that the perceived shape and motion properties of the objects were unaffected by illumination transformations these tasks were removed from the translation experiment and the subjects performed only the indistingushablity and illumination difference tasks.

The results of the translation experiment are summarized in Figure 10c,d. Under the conditions tested, the warp transformation (W) produced only indistinquishability or equivalence (not illustrated), and the pixelwise interpolation transformation (P) produced almost only non-equality. The most interesting behavior was exhibited by the reprojection transformation (RP) which showed equivalence at small keyframe intervals (k=10), mostly non-equality at large intervals (k=50), and mixed behavior in-between (k=25), however one counterintuitive result was that at this level, equivalence increased with surface gloss (M3 to M0) opposite to the result reported by [4] We speculate that it may be that the dynamic pattern of glints produced by the glossy moving object is so visually complex that it makes it hard to distinguish between the reference and transformed renderings, but further investigation of this issue is underway.

4. CONCLUSION

The main thesis of this paper is that imaging systems research and development might benefit from a different perspective on image quality. Basic image signal characteristics like intensity, contrast, noise, color gamut, spatial resolution, and temporal response are all important contributors to image quality, but they are not the whole story. Realistic images contain information about objects and scenes that are encoded in the patterns of light the images send to our eyes. The illustrations and experimental findings presented in this paper are intended to show that these patterns and the visual information they represent may be robust to both conventional image distortions such as contrast reduction and noise, and novel distortions such as warping. This insight leads to the concept of visual equivalence: that visibly different images (with possibly different conventional levels of quality) may be equivalent as visual representations of object and scene properties. Not to disrespect McLuhan, but regarding image quality, "The medium is *not* the message", and the signal and visual information properties of images need to be distinguished. The work presented on visual equivalence metrics and predictors represents some first steps in this direction. The hope is that the newfound latitude provided by the equivalence perspective may allow the development of efficient, high fidelity image capture, coding, rendering, and display algorithms that push beyond current signal-centric concepts of image quality.

ACKNOWLEDGMENTS

This work was supported by the Chester F. Carlson Center for Imaging Science, the RIT Office of the Vice President for Research, and the National Science Foundation (NSF/CPA-0811680).

REFERENCES

1. McLuhan, M. (1964) *Understanding Media*, McGraw-Hill, New York.
2. http://www.arizona-software.ch/audioxplorer/
3. http://ugandabeat.files.wordpress.com/2009/09/mustang.jpg
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. Ferwerda, J.A., Ramanarayanan, G., Bala, K. and Walter, B.J (2008) Visual Equivalence: an object-based approach to image quality. Proceedings IS&T 16th Color Imaging Conference, 347-354.
6. VanGorp, P., Condon, T., Ferwerda, J.A., Bala, K., Schoukens, R., and Dutre, P. (2009) Visual equivalence in dynamic scenes. Department of Computer Science, Katholieke Universiteit Leuven, Tech. Rep. CW 557
7. 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.
8. Lubin, J. (1995) "A visual discrimination model for image system design and evaluation. In *Visual Models for Target Detection and Recognition*, E. Peli, Ed., World Scientific Publishers, Singapore, 207–220

9. 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.
10. Debevec, P. and Malik, J. (1997) Recovering high dynamic range radiance maps from photographs. Proceedings SIGGRAPH ’97.
11. Fleming, R.W., Dror, R.O., and Adelson, E.H. (2003) Real-world illumination and the perception of surface reflectance properties. Journal of Vision 3(5), 347–368.
12. Ward, G.J. (1992) Measuring and modeling anisotropic reflection. Proceedings SIGGRAPH ’92, 265-272.
13. Dror, R.O., Willsky, A.S., and Adelson (2004) Statistical characterization of real-world illumination. Journal of Vision 4, 9, 821–837.
14. Motoyoshi, I., Nishida, S., Sharan, L. and Adelson, E. H. (2007) Image statistics and the perception of surface qualities. Nature 447, 206–209.
15. Sharan, L., Yi, Y., Motoyoshi, I., Nishida, S., and Adelson, E.H. (2008) Image statistics for surface reflectance perception. J. Opt. Soc. Am. A, 25(4), 846-865.