Journal of Imaging Science and Technology® 55(3): 030505-1–030505-13, 2011.

Effect of Scene Content on the Perceptibility of Differential Gloss

Susan P. Farnand▲
Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, New York 14623-5603
E-mail: farnand@cis.rit.edu

Abstract. *Differential gloss is the term used to describe the condition where areas of a printed image, especially adjacent areas, appear to reflect light in different ways giving these areas distinctly different gloss appearances. This phenomenon is quite common in dry toner electrophotographic imaging and some ink jet imaging technologies. Differential gloss, while well known, is difficult to quantify in a meaningful way. One of the tasks undertaken by international standards community was to develop an image quality scale for the visual attribute of differential gloss. In an attempt to do this, experimentation was conducted using prints of three scenes. The results of this experimentation showed that the rankings made by the observers were scene dependent, indicating that the single number provided by measuring gloss differences on a patch target would be insufficient to describe the differential gloss perceived in complex images. The experimentation described in this study was undertaken to examine the effect of scene content on the perceptibility of differential gloss. The results indicate that gloss differences are more perceptible when they are central to the scene, are separated by well-defined edges, and of a feature size greater than about 0.5 cm but less than about 10 cm. The experiment also provided evidence that the presence of a human face can increase the perceptibility of differential gloss. © 2011 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2011.55.3.030505]

INTRODUCTION

Differential gloss is a phenomenon where different areas of the same image reflect light differently, leading to an uneven appearance of gloss across the image. This phenomenon is quite common in dry toner electrophotographic imaging and some ink jet imaging technologies where the first surface reflection properties of the toner or ink and the substrate can be markedly different. Also, high density areas composed of multiple layers of toner can have substantially different specular reflection properties than low density areas composed of a sparse layer of toner through which areas of substrate remain visible. Differential gloss can produce effects in complex images such that detailed areas have almost the impasto appearance of an oil painting or that large uniform areas are highly reflective while adjoining areas are not. These effects can be pleasing or disturbing, depending on the image, the application, and the observer. Depending on the application, differential gloss may be something to avoid or accentuate. In either case, having an understanding of the parameters governing the perceptibility of differential gloss is of interest. Being able to measure perceptibility of differential gloss would be useful in the development of printers and printer products such as inks, toners, and substrates. Developing a perceptual image quality scale for differential gloss is one of the objectives of the International Committee for Imaging Technology Standards (INCITS) W1.1 Image Quality for Printer Systems *ad hoc* committee established by W1, the Office Equipment Subcommittee of INCITS that is the ANSI Technical Advisory Group for the ISO/IEC Joint Technical Committee, which is responsible for the standardization of the arena of Information Technology.[1] Toward this end, members of the committee conducted experimentation in 2007 using prints of three scenes made on equipment exhibiting a range of differential gloss behavior.[2] The prints were visually scaled by observers in several locations across the United States. To make comparative objective measurements, prints of the patch target developed by the W1.1 Committee on Perceptual Measurement of Gloss were also made. A single patch target was being used to quantify the level of differential gloss exhibited by a given printer. The results of this experimentation indicated that, for the printers exhibiting a high degree of differential gloss, scene content had an impact on how the prints were scaled. This suggests that the single number provided by measuring gloss differences on a patch target would be insufficient to describe the relative differential gloss perceived in complex images. The experimentation reported here was undertaken to explore the effect of scene content on the perceptibility of differential gloss.

BACKGROUND

For the past decade, a team has been in place with the charter to develop perceptual image quality metrics. The INCITS W1.1 Image Quality for Printer Systems *ad hoc* committee was established by W1, the Office Equipment subcommittee of INCITS that serves as the ANSI Technical Advisory Group for ISO/IEC Joint Technical Committee 1, which is responsible for the standardization in the arena of Information Technology.[1] One of the tasks undertaken by this committee is to develop image quality measures for the visual attributes of gloss and gloss uniformity.[2,3] Gloss is defined as a surface attribute that makes it appear shiny or lustrous and

▲IS&T Member.

Received Aug. 23, 2010; accepted for publication Mar. 5, 2011; published online Apr. 29, 2011.
1062-3701/2011/55(3)/030505/13/$20.00.

Figure 1. The ISO/IEC 19799 differential gloss test chart (Source: Farnand and Nilosek, "Measurement of Differential Gloss Using a μ-Goniophotometer").

is generally related to specular reflectance.[4] Gloss level is typically determined by measuring the specular reflectance of a surface at one of the several specified measurement angles, though some researchers have found better correlations between measured and visually scaled data following other approaches.[5–7] Research has also been conducted and models were developed for realistic rendering of apparent gloss of displayed images.[8–10]

In contrast to gloss level, which is a characteristic of a surface, differential gloss is the phenomenon where individual segments of a printed image exhibit varied levels of gloss and is generally considered an artifact to be avoided. The initial work by the W1.1 standards *ad hoc* committee on gloss and gloss uniformity toward a perceptual metric for differential gloss, as opposed to gloss level, was conducted with the ISO/IEC Differential Gloss test target (Figure 1), rather than pictorial images, to specifically "avoid the image content influence."[2] In subsequent experimentation conducted by this committee, observers rated prints of three complex images made on a variety of printers for their level of apparent differential gloss. In Figure 2, the visual scaling values resulting from this testing are plotted relative to the natural logarithm of the maximum range of the gloss measurements made of the 40 patches on the ISO/IEC Differential Gloss test target (Fig. 1). For a given printer, which has one value for measured differential gloss, it was found that prints having different scene contents could receive statistically significantly different visual differential gloss ratings. This can lead to the situation where prints of one scene can receive similar visual ratings for differential gloss for two different printers that have significantly different measured differential glosses while prints of another scene can receive visual ratings that are similar and measured values that are substantially different. Note, for example, the "picnic" scene relative to the "objects" scene for the printer having a measured gloss value of about three and the printer having the highest measured gloss value (about 4.5) (Fig. 2). The "objects" scene adequately represents the difference in both the measured and perceived values of differential gloss for the two printers. However, there is no significant difference in the perceived differential gloss in the picnic scene for these two printers, though the measured gloss ranges on the test target are substantially different. It would be interesting, then, to understand what is different about the picnic scene relative to the object scene in terms of the image features that drive the perceptibility of differential gloss. This information would be imperative to develop test targets that may be used to reliably evaluate printers for their level of perceptual differential gloss.

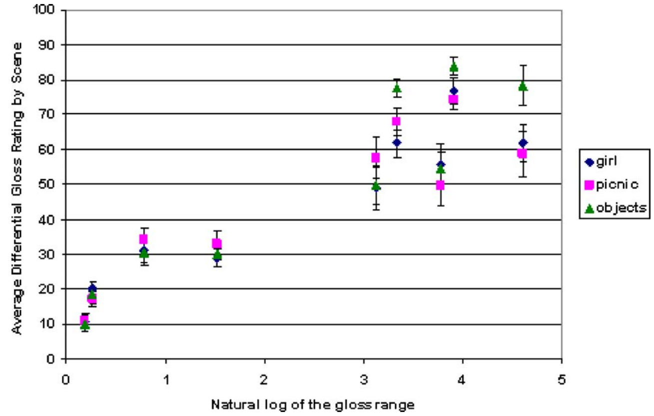

Figure 2. Mean differential gloss rating as a function of the 20° gloss range for each of the nine prints. (Source: Farnand and Nilosek, "Measurement of Differential Gloss Using a μ-Goniophotometer").

EXPERIMENTATION

The experimental procedure used in this study was the triplet comparison method described in ISO 20462-2.[11] This procedure can be thought of as a paired comparison approach in which three pairs are evaluated simultaneously rather than one or as a rank order approach in which the stimuli are ranked three at a time. The results of this procedure are translated from rankings to scale values in the same way as for either the rank order or paired comparison methods using an approach based on case V of Thurstone's law of comparative judgments.[12–14] Specifically, each image that was identified as having a more perceptible level of differential gloss received a "1." The total number of times a given image was identified as having more apparent differential gloss than other images was tallied and this sum was divided by the total number of comparisons made for each image (the number of observers). These values were translated to Z scores using standard tables.

Figure 3. The experimental setup. Half of the bulbs in the outer two sets of lights were illuminated. The remaining lights were turned off.

The experiment was conducted under simulated D50 lighting conditions in a viewing room in the Color Science building at the Rochester Institute of Technology. The lighting configuration was designed such that light was coming from an angle and not directly overhead; the lights directly overhead were turned off, while the lights to the left and right were illuminated (Figure 3). Previous testing indicates that the results would be similar with direct lighting but that this design made the differential gloss easier to see, making the task easier for observers.[2,7]

The observers were seated in front of the viewing table and were allowed to handle the prints. The viewing distance was not held constant but was typically maintained and reading distance. The observers were given a brief explanation of the phenomenon of differential gloss as the difference in gloss between the high gloss and low gloss areas within a single image and shown an image in which the gloss difference was readily apparent. The observers were instructed to order the prints from the one in which the difference in gloss was easiest to see to the one in which the differential gloss was hardest to see. The stimuli were presented three at a time. The triplets used were randomized for each observer.

The experiment involved two segments. The stimuli for all of the testing were 4×6 in.^{2} prints produced on an electrophotographic printer in the Digital Printing Center in the College of Imaging Arts and Sciences at RIT on uncoated paper and mounted on 6×8 in.^{2} midgray card stock. In the first segment, the stimuli comprised patch images in which the patches were of varying sizes (Figure 4). Consistent colors were used for the patches in these images so that color and differential gloss level would be held constant. Differences in appearance of differential gloss must then be due to

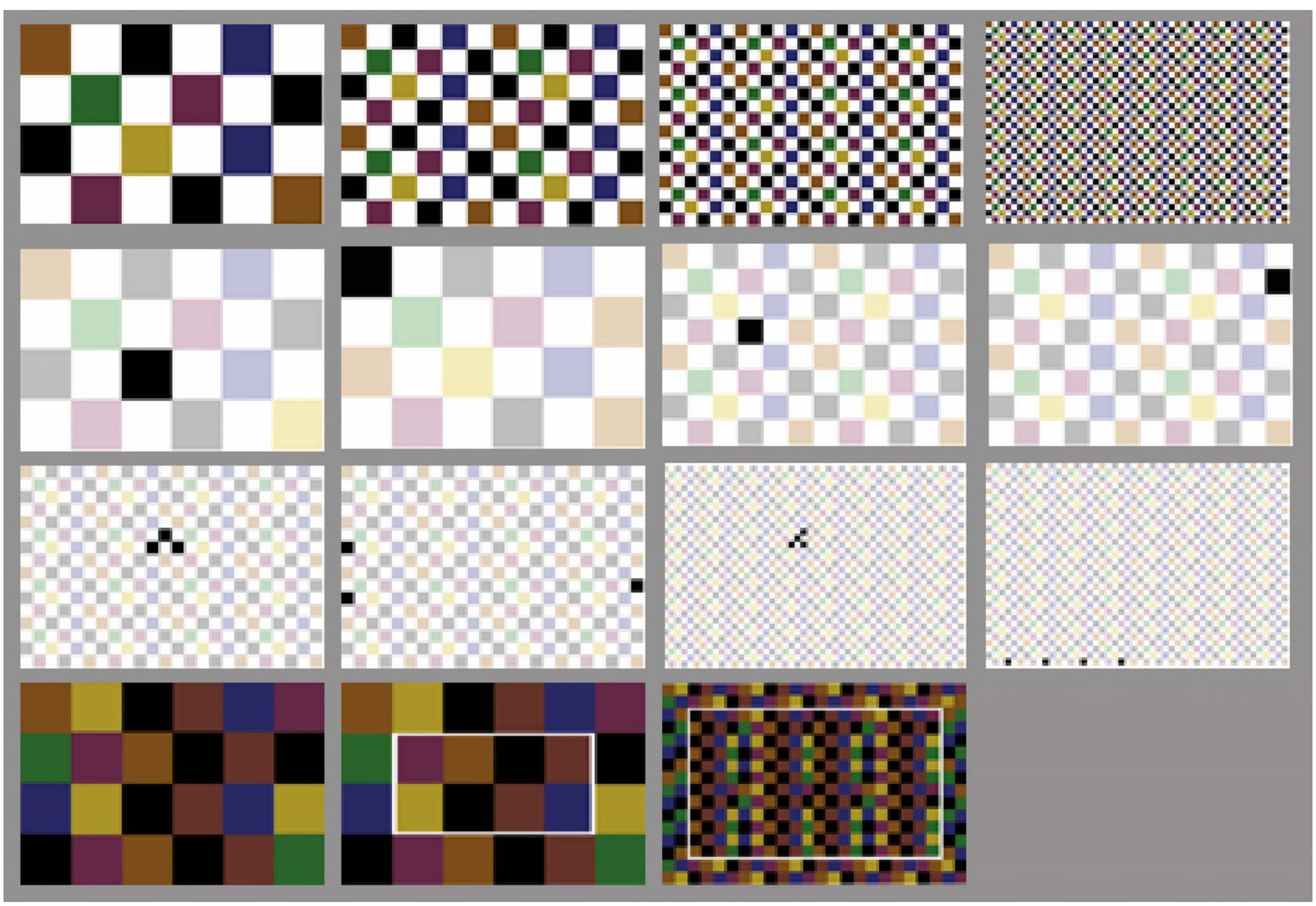

Figure 4. The patch images used in the experiment. From top left, they are referred to as big squares, half squares, quarter, eighth, big mid, big side, half mid, half side, quart mid, quart side, eighth mid, eighth side, saturated, line mid, and line side.

Figure 5. The pictorial images used in the experiment. From top left, they are referred to as Park bench, kids, hockey, lute, leaves, falls, paradox, swamp, bahnhof, rose, candles, flowers, palace, cooking, and Tibetans.

patch size and position. The one exception was the images, "saturated" in which all of the patches were highly saturated with many black patches and no white (substrate) areas. The other images included two images in which there is a thin line of bare substrate, a quarter to a third of the way into the image, in some images half of the patches were highly saturated and the other half were the substrate, and in the remaining images, only one to three patches, depending on patch size, were highly saturated. In the images having only a few highly saturated patches, some had these patches near the center of the image and some had them at the edge of the image. Due to the electrophotographic process and the uncoated media used, the highest density areas of the images had the highest measured gloss [the black patches had a measured gloss value of 52.8 gloss units (GUs) at 60°] and the white (substrate) areas had the lowest gloss (5.1 GU at 60°), resulting in a differential gloss value of 47.7 for all images except the saturated image, which had no visible substrate. This image had a minimum measured gloss of 22.9 GU at 60° and, therefore, a differential gloss of about 30 GU.

In the second segment of the experiment, the images were pictorial scenes in landscape format (Figure 5). These images were again printed on the same printer and substrate. Most images contained some areas of both white and black, resulting in differential gloss comparable to the patch images. Exceptions were the "paradox" image, which had a minimum gloss level of 6.8 GU at 60° (differential gloss of 46 GU) and the "kids" image, which had a minimum gloss level of 8.3 GU at 60° (differential gloss of 44.5 GU). The "bahnhof" and "flowers" images had limited areas of white but also had areas of yellow that had lower gloss levels (about 13 GU) that were included in the low gloss areas in the analysis.

Fifteen observers participated in the experiment. These observers included seven females and eight males. Nine of the observers were in their early 20s; six of the observers were in their 30s or 40s. Following the experiment, a confirmation run was conducted that also used pictorial imagery.

RESULTS AND DISCUSSION

Patch Image Experiment

The results for the patch images used in the differential gloss scaling experiment are shown in Figures 6–8. The graphs in Figs. 6 and 7 show the visual ranking values relative to the mean ranking calculated for two specific groups of observers. Fig. 6 shows the results for the main group of ten observers, while Fig. 7 shows the results for the remaining five observers. These two groups are each composed of observers that were in close agreement with one another, each having a mean correlation with the group mean of around 0.9. (This value was calculated by finding the mean rank value for each image across observers within the group, calculating the cor-

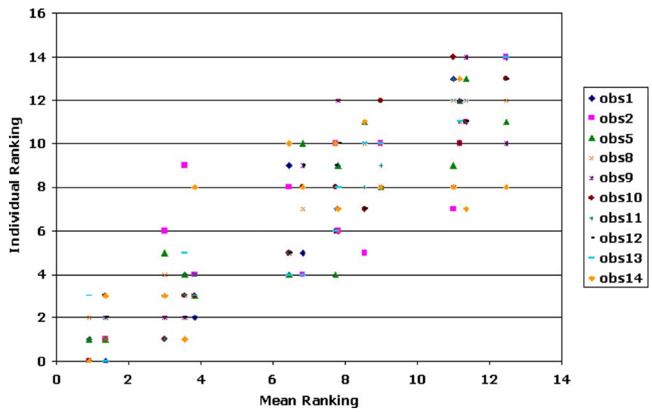

Figure 6. The visual rankings for each of ten observers relative to the mean rank values for those observers for the patch images.

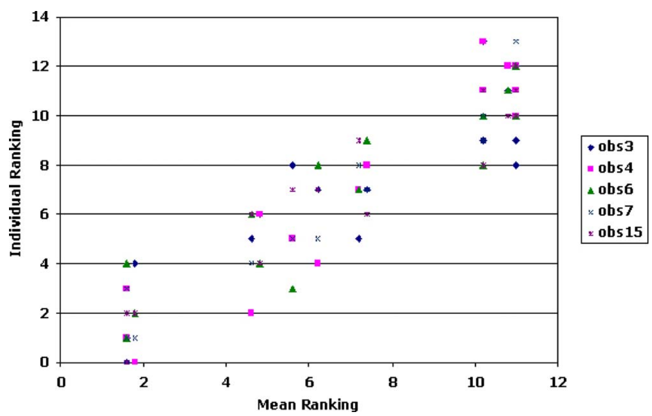

Figure 7. The visual rankings for the remaining five observers relative to the mean rankings for those observers for the patch images.

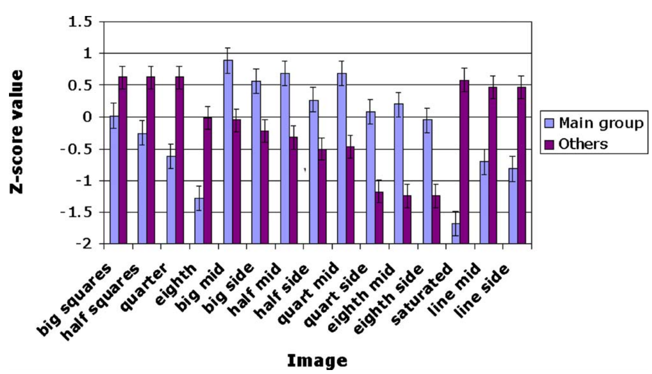

Figure 8. The Z-score values by image for the two groups of observers in the experiment for the patch images.

relation coefficient between each observer and that mean and then taking the mean of those correlation coefficients.) These two groups, however, are highly negatively correlated with one another. The graph in Fig. 8 show the mean Z-score values determined for each image for each of the two groups. Clearly, there are significant differences for the two groups of observers. From inspection of Fig. 8, it is readily apparent that while the larger group of observers found the images having a single or a small number of highly saturated higher gloss patches to have the higher level of differential gloss perceptibility, the smaller group ranked the images having half the patches or more highly saturated and higher overall gloss to have more perceptible differential gloss.

There are several reasons why this result might have occurred. First, the smaller group may have misunderstood the task and may have been scaling overall gloss level rather than the difference in gloss. Evidence that this may have been the case is that the image having fully saturated patches with no areas of substrate visible was rated as one of the images having the most perceptible differential gloss despite the fact that, as measured by a glossmeter, the level of gloss across the image is relatively even and high, with a minimum gloss of about 23 GU at 60° and a differential gloss value of about 30 GU, which was substantially lower than the 47.7 GU value for the other images. There is, however, a fair degree of gloss texture, which produces an "orange peel" appearance in the high gloss areas of the images that may have caught the attention of this group of observers. This textured appearance may have been more apparent to observers having higher visual acuity. The acuity of the observers in this experiment was not tested. However, the five observers in the smaller group were all in their early 20s. It may be that, for this group of observers, having more edges demarcating high and low gloss areas actually did make the differential gloss more apparent than having one or a few discrete areas that were significantly different from the rest of the image.

It is also true that four of the five observers in the smaller group were among the first to participate in the experiment. It may be that the administrator became more adept at describing the differential gloss artifact under scrutiny as the experiment progressed. (Although the observers were verbally given the same instructions, they were given the opportunity to ask questions and some required further clarification of the differential gloss artifact.)

Along with the differences in the results between the two groups of observers, there were some important similarities as well. First, both groups found it easier to see differential gloss with larger squares relative to smaller squares. For the main group, the differential gloss perceptibility rating dropped steadily with the size of the squares, especially for the images having saturated squares (light blue bars in Fig. 8), though these declines were not always statistically significant. For the second group of observers (dark purple bars in Fig. 8), the differential gloss perceptibility ratings were not statistically significantly different for the images having 1, 0.5, and 0.25 in. saturated squares. However, the mean rating for the image having 8 in. squares was significantly lower. Furthermore, the ratings did tend to drop with size when there was a single or a few saturated squares in the middle of the image or on the side of the image, though the decreases were not always statistically significant.

The decreasing differential gloss perceptibility with the size of the squares may be explained by the idea that, as the areas that were different in gloss content became smaller, the

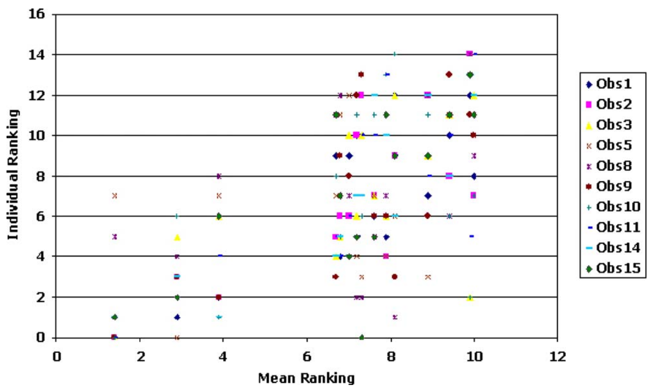

Figure 9. The visual rankings for each of ten observers relative to the mean ranking for those observers for the pictorial images.

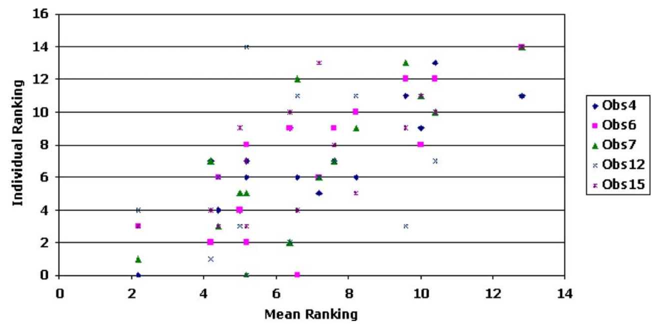

Figure 10. The visual rankings for the remaining five observers relative to the mean ranking for those observers for the pictorial images.

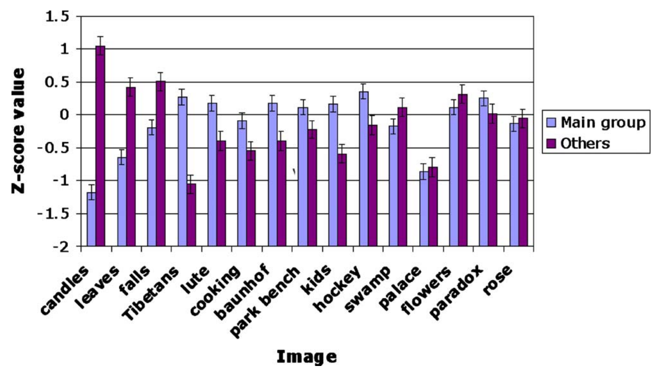

Figure 11. The Z-score values by image for the two groups of observers for the pictorial images.

individual gloss squares become increasingly difficult for the visual system to resolve and the gloss would tend to blend together giving the image a relatively uniform gloss appearance. It also makes sense that the main group, which included all of the older observers, might include observers for whom this inability to resolve the gloss differences might occur for larger squares than for other, younger, observers.

Another result shared by all observers was that images having small areas of high gloss had higher ratings for the perceptibility of differential gloss when these areas were in the center of the image than at the edge (Fig. 8). This result is in agreement with research suggesting that humans tend to look at the central area of an image first before looking at its edges.[15] Once again, this result is more distinct for the main group of observers than for second group, which only had a statistically significant result for the 0.25 in. squares. It may be interesting to note that the 0.25 in. squares had three high gloss squares located at different edges of the image, unlike the 8 in. squares, which had four squares all in close proximity to each other. It was not possible, at the viewing distances used, to simultaneously have all three squares within the foveal region of view for the image with 0.25 in. high gloss squares at the edges, while it would have been possible with the image having 8 in. squares. This may explain why there was the most significant difference between image having high gloss squares in the middle versus the edges for the images having 0.25 in. squares for both groups of observers.

Pictorial Image Experiment

The results for the images having complex content are shown in Figures 9–11, with the visual ranking relative to the mean ranking for two groups of observers shown in Figs. 9 and 10 and the Z-score values by image shown in Fig. 11. The results for this set of images are noisier, as is evident from the greater spread of the data in Figs. 9 and 10 relative to Figs. 6 and 7, with mean correlation with the mean rank value being about 0.65 for the main group of observers and about 0.8 for the second group of five observers who agree better with one another than with the main group.

The "candles" image had the most distinctly different results for the pictorial image set with the main group of observers finding it to have the lowest apparent differential gloss while the second group found it to have the highest. This image is largely black, and, therefore, of high overall gloss level with four small regions, the candle flames, where substrate is visible. In both composition and result, this image was most similar to the saturated and "eighth" images in the patch image set. The "leaves" and "falls" images have compositional similarities to these images as well. The leaves image is also largely black; however, it has a larger area where the substrate is visible. The falls image has a significant black region but could not be characterized as mostly black and has an area of white, where substrate would be visible, similar to the leaves image.

In comparison, the image that was ranked high for differential gloss by the main group but was identified as having the lowest level of differential gloss perceptibility by the second group was the "Tibetans" image. This image had a large dark high gloss area along the bottom, while the rest of the image was relatively matte, making it closest in composition to the "big side" image (Fig. 4). Again the results for these two images are basically in agreement, although the difference in rankings for the Tibetans image was larger than for big side.

The other images that were all identified as having a high level of differential gloss perceptibility by the main group but relatively low perceptual differential gloss by the second group of observers were the "lute," "cooking," "bahnhof," "park bench," "kids," and "hockey" images. The first four of these images had a mid- to high level of gloss overall with one distinct central region of lower gloss, while the "kids" and "hockey" images had the opposite composition—an overall low level of gloss with a few distinct regions of high gloss near the center of the image. The central region that differed in gloss appearance in the "lute," "cooking," "kids," and "hockey" images was relatively large—on the order of 1 in. or so. These four images most closely resembled the "big mid" patch image in composition. The region in the "bahnhof" and "park bench" images that differed in gloss was smaller, approximately 0.5 in., making these images more like the "half mid" image. The big mid and half mid images were also ranked higher for differential gloss perceptibly by the main group than by the second group.

It may be worth noting that the main group of observers rated all of the images containing people, "kids," "lute," "cooking," and "Tibetans," relatively high for differential gloss. It is reasonable to assume that the presence of people in an image would affect the acceptability of differential gloss. Research indicates that observers are drawn to human faces more than any other scene content; that people look first and longest at these image regions.[15] In this experiment, one observer commented that he found the differential gloss more disturbing when it occurred around people's faces. It is possible that having people in images affects perceptibility as well by increasing the level of attention to areas where the gloss differential is present. Though these images were similar in composition to patch images that were also ranked higher by this group, it is possible that this difference in image content was also a factor. It may be the inclusion of people that pushed the difference in the rankings between the two groups for the "Tibetans" image to be greater than might be expected from the size and location of the picture elements based on the data from the patch targets.

Observers were in statistical agreement in their rankings of the remaining images, "swamp," "palace," "flowers," "paradox," and "rose." The "swamp" image was actually rated just slightly higher by the second group than the main group. The "swamp" image featured light-colored low gloss trees among darker glossier foliage, creating thin stripes of low gloss areas, making it something like a cross between the "line mid" (due to the thin stripes) and half mid (due to limited number of areas that differed in gloss) patch images.

None of the observers saw much in the way of differential gloss in the "palace" image. Both groups of observers ranked it the second lowest for differential gloss perceptibility. This image has a thin line of low gloss area down the left side and had a few other low gloss areas interspersed within it. However, none of these regions is particularly distinct. In composition, then, it is probably closest to "half side" or "quart side."

The "flowers," "paradox," and "rose" images were all ranked high for differential gloss by both groups of observers. However, they each had different general compositions. The "flowers" image had high overall gloss with a several low gloss areas; the signs, which were each about 0.5 in. square, scattered throughout the image. It should be noted that these signs were yellow, not white, so the measured differential gloss is lower than it would be if they were white. This image was closest in composition and result to something between the "half squares" and half mid patch images because there were several areas having a distinctly different gloss level not just one as in half mid but not nearly as many as in half squares.

The "paradox" image, which is an image of a mountain lake, is basically half high gloss and half low gloss with the demarcation line essentially coinciding with the line where the hills and trees meet the sky and the water through the center of the image. This image did not contain any areas that were basically white; the sky and water areas had the lowest gloss (6.7 GU at 60°), meaning the differential gloss was about 45 GU rather than the 47 GU at 60° differential gloss level for other images, yet this image was consistently ranked by all observers as having a high level of perceived differential gloss. This scene did not have a corresponding image relative to its composition among the patch images.

Finally, the "rose" image had a single large central area of low gloss. This area was considerably larger than the 1 in. square of big mid. It was found with the patch image set that the perceptibility of differential gloss generally decreases with patch size. Given this, it might be expected that the rose image with its 2.5–3 in. diameter area of low gloss embedded in high gloss foliage should be ranked among those having the most apparent differential gloss, especially by the main group. This is not the case. It may be that when the area of the image having the difference in gloss increases past a certain point, it becomes larger than the foveal region of vision, meaning that not all of the edges of this area can be fixated simultaneously. We might be able to consider this image then as being, much like the "swamp" image, a cross between the line mid and big mid patch images. Indeed, the Z scores for the "rose" and "swamp" images are statistically in agreement.

MODEL DEVELOPMENT

The experimental results point to several factors that might be important in the perceptibility of differential gloss in complex images. The results for the patch images indicated that the size and location of the higher (or lower) gloss areas in the image may be important. In the experiment, images having a high gloss patch (or patches) of about 1–3 cm in the central area of the image were ranked most highly for perceptibility of differential gloss. The rankings for pictorial images such as "bahnhof," "flowers," and "park bench" reinforce this finding. The results for the pictorial images further indicated that images involving a human face might receive a boost to the perceptibility of their gloss differences.

To explore the possibility that these factors could be combined to create a model for predicting the perceptibility of differential gloss, the pictorial images were processed to isolate regions or boundaries of high gloss difference. To generate objective values for the size and location of the gloss areas, the pictorial images were first down sampled in Adobe Photoshop® to 432 pixels wide by 288 pixels high

Table I. Processing results for the pictorial images included in the experiment.

	Areas		Pixels		Size		Edges		Edge			
Image	Central	All	Center	All	Center	All	Center	All	Combo	Quality	Contrast	Face factor
Park bench	1	14	1,044	5,103	0.20	0.07	2.2	8.8	5.5	0.5	0.6	2
Kids	2	5	9,735	14,848	0.94	0.57	6.3	8.5	7.4	0.3	0.48	3
Hockey	4	12	2,407	8,535	0.12	0.14	8.6	8.6	8.6	0.6	0.78	2
Lute	1	4	2,822	4,242	0.54	0.20	6.8	10.4	8.6	0.6	0.64	3
Leaves	2	4	4,261	6,343	0.41	0.31	5.6	6.2	5.9	0.1	0.43	1
Falls	2	2	3,436	3,436	0.33	0.33	9.6	9.6	9.6	0.2	0.48	1
Paradox	1	2	16,870	28,009	3.25	2.70	10.5	10.5	10.5	0.8	0.58	1
Swamp	1	2	837	8,700	0.16	0.84	3.8	7.8	5.8	0.8	0.65	1
Bahn	1	2	2,862	3,969	0.55	0.38	4.0	5.0	4.5	0.9	0.43	1
Rose	1	2	23,638	25,243	4.56	2.43	7.7	9.9	8.8	0.4	0.7	1
Candles	4	5	1,426	1,426	0.07	0.06	3.1	3.9	3.5	0.1	0.77	1
Flowers	6	8	7,590	8,548	0.24	0.21	7.5	10.5	9.0	1	0.75	1
Palace	2	5	705	2,801	0.07	0.11	2.3	6.6	4.5	0.1	0.73	1
Cooking	1	1	5,819	5,819	1.12	1.12	6.4	6.4	6.4	0.5	0.5	2
Tibetans	1	3	1,304	1,304	0.25	0.08	6.2	7.2	6.7	0.5	0.33	3

using bicubic image interpolation. Down sampling was conducted because the experimental results indicated that smaller features did not seem to play a significant role in the appearance of gloss differences. The next step was to identify areas of the image where differences in gloss occurred. With this image set, black or dark areas had higher gloss levels and white or light areas had lower gloss levels. (This was true because the fused toner had a higher gloss level than the paper substrate.) Identifying regions of high gloss differential then required finding regions where the content shifted from black or dark colors to white or light colors.

To do this, images with limited areas of white (the “kids”, “paradox,” “bahnhof,” “flowers,” and “Tibetans” images) were first posterized at two levels to isolate areas that were light but not white. This step proved necessary because, although the differential gloss level between yellow or light cyan areas and black areas did not measure as high as between black and white (40–46 GU as compared to about 48 GU), it was high enough to be clearly visible. Since the “paradox” and “flowers” images were highly ranked for perceived gloss (Fig. 11), the decrease in measured differential gloss appeared to be less of a factor in the perceptibility of the differential gloss than other possible parameters. The posterization technique separated the images into cyan, magenta, yellow, red, green, blue, black, and white areas. (For the “paradox” image, most of the sky and the central area of the water were classified as white.) All of the images were then thresholded at an arbitrarily chosen level of 192 or three-quarters of the 256 possible levels. This process was conducted with the images in RGB mode. The thresholded images were visually evaluated for white areas central to the image, where the central area is defined as an ellipse having axes 2 in. high and 4 in. wide on the 4×6 in.^{2} prints. The number of areas, their average size, and their approximate linear edge length were recorded. The same characteristics for the white areas in the periphery were then determined. The presence of human faces in the images was noted where applicable. Also, it was speculated that the contrast between and the character of the edges between the high and low gloss regions might be of importance in determining the perceptibility of gloss differences. The contrast level and percentage of edges that were relatively straight (identified visually on the thresholded image and the linear edge length was compared to the total edge length of the white areas in the image) were consequently noted as well.

The results of this evaluation, listed in Table I, were used as input in a stepwise multiple linear regression analysis using Minitab® statistical software. The results of this analysis, included in Tables II and III and summarized in Table IV, indicate that edge characteristics including the length of the edges between high and low gloss areas in the image, especially in the central part of the image, and the edge quality or relative straightness of the edges had the strongest influence on the perceptibility of gloss differences. The other factor playing a significant role was the presence of a human face. These three factors resulted in a reasonable model for predicting perceived differential gloss. Note in Table IV that with these three factors included, the adjusted R-squared value between the predicted ranking and Z-score values was 0.85. The number of low gloss areas in the central region of the image also had a significant effect on the predictive power of the model. Moreover, the average size of the gloss features in the image and the edges in the entire image provided additional information in the model. The model developed using Minitab® for predicting how differential gloss catches the attention of observers is shown in Eq. (1),

Table II. Minitab® stepwise linear regression analysis results.

Step	1	2	3	4	5	6
Constant	−0.7323	−1.0588	−1.3381	−1.1752	−1.1138	−1.2324
Edge quality	1.32	1.25	1.09	1.11	1.16	1.11
T value	4.59	5.56	5.96	6.64	7.04	6.96
P value	0.001	0	0	0	0	0
Face factor	0.227	0.214	0.186	0.14	0.12	
T value	3.12	3.76	3.43	2.24	2	
P value	0.009	0.003	0.006	0.052	0.08	
Edges—center	0.061	0.058	0.074	0.055		
T value	2.94	3.03	3.38	2.23		
P value	0.013	0.013	0.008	0.056		
Areas—center	−0.057	−0.082	−0.092			
T value	−1.79	−2.3	−2.68			
P value	0.103	0.047	0.028			
Size—all	−0.019	−0.022				
T value	−1.36	−1.66				
P value	0.208	0.136				
Edges—all	0.04					
T value	1.46					
P value	0.183					
S	0.289	0.224	0.175	0.159	0.153	0.144
R-Sq	61.87	78.96	88.22	91.08	92.6	94.15
R-Sq(adj)	58.94	75.45	85.01	87.52	88.48	89.76
Mallows Cp	41.1	19.8	9.1	7.2	7.1	7

Table III. Analysis of variance.

Source	DF	SS	MS	F	P
Regression	6	70.386	11.731	78.17	0
Residual error	8	1.201	0.15		
Total	14	71.587			
Source	Seq SS				
Edge quality	1.776				
Face factor	0.298				
Edges—center	0.255				
Areas—center	0.267				
Size—center	0.057				
Edges—all	0.030				

$$\mathrm{dg}_a = -1.23 + 1.11e_q + .12\ \text{face} + 0.055e_c - 0.092a_c - 0.022s_{\mathrm{gf}} + 0.04e_a, \quad (1)$$

where dg_a is the apparent differential gloss scale value, e_c is the length of edges in the central part of the image, e_q is the edge quality represented by the proportion of the edges between high and low gloss areas that are straight, face signals the presence or absence of a human face in the image, a_c is

Table IV. Results of Minitab® stepwise linear regression analysis.

Predictor	Coefficient	SE coefficient	*T*	*P*	Step	*R*-Sq(adj)
Constant	−1.232,4	0.178,8	−6.89	0		
Edge quality	1.105,6	0.158,8	6.96	0	1	59
Face factor	0.054,8	0.024,57	2.23	0.056	2	75
Edges—center	0.120,4	0.060,21	2	0.08	3	85
Areas—center	−0.092,0	0.034,37	−2.68	0.028	4	88
Size—all	−0.022,1	0.013,34	−1.66	0.136	5	88
Edges—all	0.040,1	0.027,5	1.46	0.183	6	90

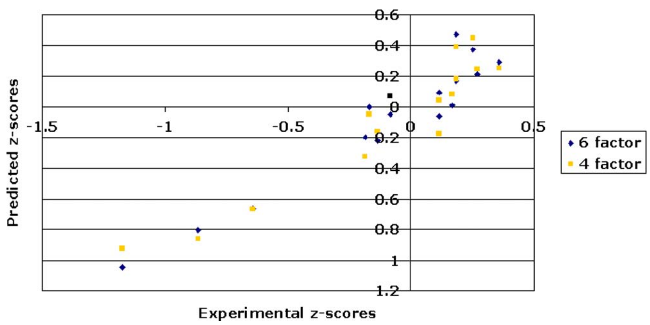

Figure 12. Predicted Z-score values using the models developed for four and six factors relative to the Z scores from the experiment.

the number of gloss features in the central part of the image, s_{gf} is the size of the gloss features in the image, and e_a is the length of edges in the overall image. The Z scores determined in the experiment are shown plotted against the predicted rankings, calculated using this equation in Figure 12. This graph also shows the results if a model including only the first four factors is used [Eq. (2)],

$$\mathrm{dg}_a = -1.18 + 1.11e_q + .19\ \ \mathrm{face} + 0.058e_c - 0.057a_c. \tag{2}$$

A confirmation test was conducted to evaluate the efficacy of the developed model. In this confirmation test 11 images, shown in Fig. 13, were "processed" using the same procedure as outlined for the experimental images; the images were posterized (if low contrast), thresholded, and evaluated for feature area size, number, edge length, and other characteristics. It was found in this case that two of the images were relatively high-key meaning that there were a few high

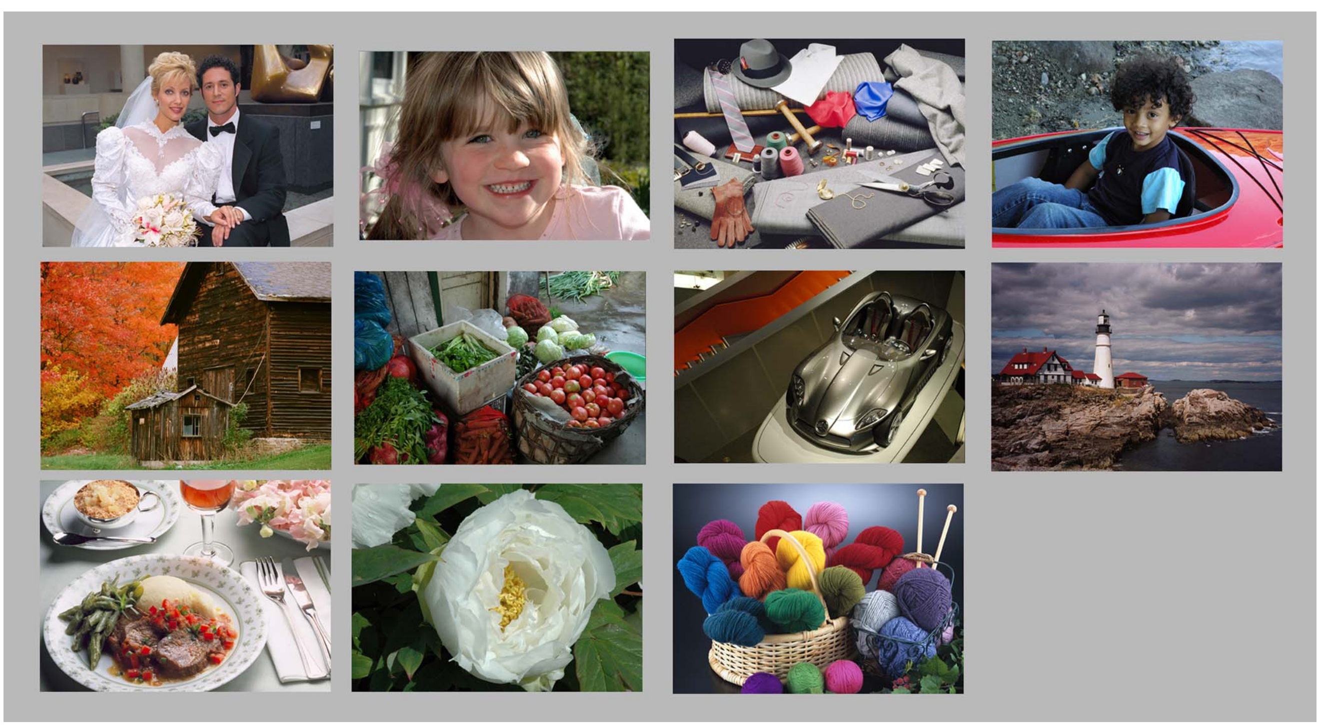

Figure 13. The pictorial images used in the confirmation test. From top left, they are referred to as bride and groom, Sarah, textiles, Kira, barns, veggies, Mercedes, lighthouse, meal, peony, and yarn. The bride and groom, textiles, barns, lighthouse, meal, and yarn are from the standard color image data (SCID) image set. The Peony image is from the RIT image set.

Table V. Processing results for the images included in the confirmation run.

Image	Areas		Pixels		Size		Edges		Edge quality	Face factor
	Central	All	Center	All	Center	All	Center	All		
Bride	2	4	39,704	40,668	3.83	1.96	13.1	18.7	0.4	3
Sarah	2	3	30,370	39,427	2.93	2.54	6.3	20.4	0.3	2
Textiles	1	2	8,040	25,363	1.55	2.45	13.6	18.2	0.6	1
Kira	3	7	6,892	9,713	0.44	0.27	9.2	18.4	0.7	3
Barn	1	3	2,168	10,870	0.42	0.57	4.2	5	0.6	1
Veggies	1	4	1,059	8,626	0.20	0.42	8.4	13.9	0.2	1
Mercedes	2	5	29,682	34,144	2.86	1.32	12.8	18.6	0.4	1
Lighthouse	1	5	1,999	6,112	0.39	0.24	5.2	14.4	0.7	1
Meal	1	3	5,100	22,344	0.98	1.44	7	14	0.2	1
Peony	1	2	26,729	28,043	5.16	2.70	13.8	21.3	0.4	1
Yarn	2	5	12,154	34,570	1.17	1.33	9	18	0.3	1

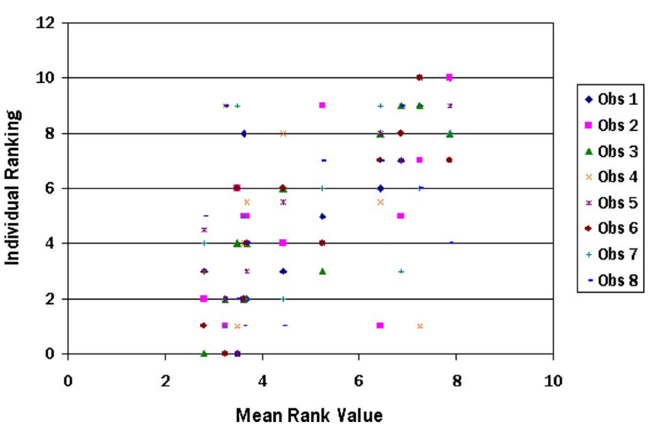

Figure 14. The visual rankings for each of eight observers relative to the mean ranking for the observers for the confirmation images.

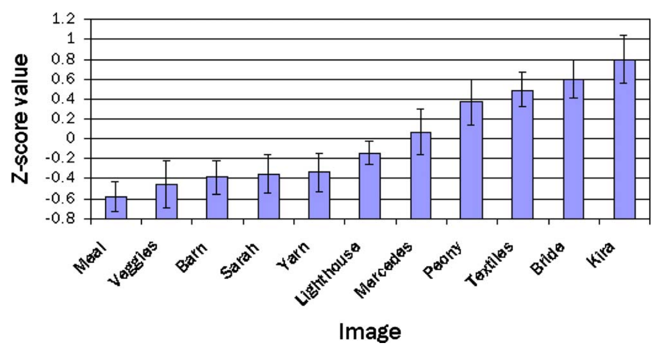

Figure 15. The Z-score values by image for the confirmation run.

gloss areas amidst a low gloss field, similar to some of the patch images in the experiment rather than a few low gloss areas in a high gloss field as was the case for the pictorial images in the experiment. For these two images, thresholding was, therefore, done at 64 rather than 192 to more adequately isolate the features of interest. The results from this process, listed in Table V, were run through the model to predict the relative perceptibility of the gloss differences in these images.

With the predictions made, a confirmation test was conducted following the same protocol as used in the original experiment. Eight observers participated, three female and five male ranging in age from about 25 to about 65. The results of the test are shown in Figures 14 and 15. Looking at the visual rankings for all of the images in Fig. 14 it is evident that most of the individual rankings fall close to the mean rankings. There were, however, two observers who each ranked one image quite differently from the other observers and three observers who ranked two images differently. One image that received outlying rankings from two observers was the "veggies" image. The tomatoes in this image had lower gloss then their surroundings, resulting in a multitude of smaller low gloss areas, somewhat off to the side. If this area happens to catch the viewer's eye, then this image may be ranked higher than if it had not.

Two of the five observers that had differing results in the initial experiment were included in the confirmation run. Since one of the observers was one of those that rated the "veggies" image differently, it may be that, for observers who produce different results from the mean rankings, having more edges demarcating high and low gloss areas actually did make the differential gloss more apparent than having one or a few discrete areas that were significantly different from the rest of the image. Alternatively, it may simply be that different things caught the eyes of these observers. Generally, the results look reasonable and no data were excluded from the analysis.

Figure 15 shows the Z-score values for all of the test images. These show that, generally, "Kira" was the highest ranked image followed by "bride," "textiles," and "peony," which were also highly ranked. The "Mercedes" and "lighthouse" images were ranked in the middle. "Sarah," "barn," "yarn," and "veggies" received lower rankings with the "meal" images tending to receive the lowest rankings of the

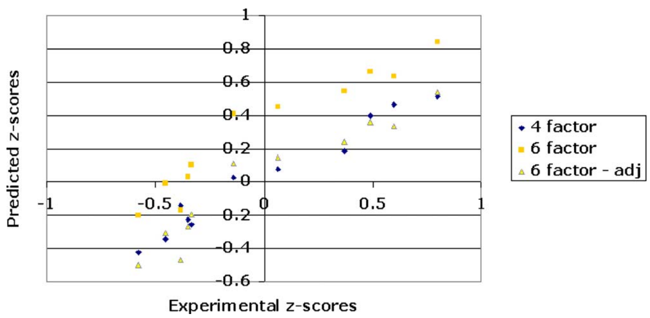

Figure 16. The Z-score values vs predicted values for the confirmation run using four, six, and six adjusted factors for prediction.

11 images. The error bars indicate, however, that the level of experimental noise is such that there is substantial overlap between these rankings, suggesting that slightly different ordering might result with additional data.

The Z-score values for the images are shown relative to the predicted Z scores in Figure 16. The results are surprisingly good, considering the level of sophistication used to determine the values for the factors feeding into the model development. The results indicate that four factors are sufficient to predict the experimental results. The two additional factors, if anything, resulted in slightly lower accuracy in predicting the experimental results.

It is possible that, with more rigorous determination of the values of the image characteristics, even better predictions could be realized; however, the agreement achieved here is sufficient for the purposes of this study. The objective of this study was to identify image characteristics having an important impact on the perceptibility of differential gloss rather than necessarily developing a model for reliably predicting it. The idea that the developed model was able adequately to predict the relative perceptibility of differential gloss in a second set of pictorial images suggests that the factors used in the model were indeed important in the perception of differential gloss. The results support the concept that the edges between high and low gloss areas, both the length of these edges, especially in the central regions of the image, and their relative straightness, have the largest impact on differential gloss perceptibility. The results also suggest that the presence of human faces has a significant positive impact on the perceptibility of differential gloss. Moreover, the results suggest that the number of significantly higher or lower gloss areas in an image influence the perceptibility of differential gloss in that image. Results also indicate that size may have an influence on differential gloss perceptibility. The results from the patch targets certainly suggested that size was relevant, though it was not needed in the model to predict differential gloss appearance in the pictorial images in the verification test. All of these factors should be considered when developing image sets for analyzing gloss appearance.

CONCLUSIONS

Experimentation was conducted to evaluate the effect of scene content on the perceptibility of differential gloss. The model generated in this study functioned well in predicting the mean rankings of the pictorial images in a confirmation run. However, it was generated using a specific printer, paper, and toner combination. Its efficacy may not hold if any of these components were changed. Different substrates and marking technologies would impact what colors appear higher in gloss level and which would appear to be lower in gloss. These changes would have to be accounted for in generating the input to the model. A different approach may be needed to properly identify the edges between high gloss and low gloss regions and a method of identifying “relatively straight” edges would be required. Having properly identified the high gloss and low gloss regions, it is possible that the model will perform adequately if the important factors for determining the effect of scene content on differential gloss have indeed been captured. It might be interesting to evaluate the model performance using images made on an ink jet printer that exhibits a relatively high degree of differential gloss where the substrate is the high gloss area and the regions having heavy ink coverage are the low gloss areas.

The goal of this study was not to develop a model for further application but to conduct exploratory work to identify factors important in the perception of differential gloss. The result that there were significant differences in the differential gloss rankings of the patch images, which all had the same differential gloss level (except the saturated image), indicates that factors in addition to differential gloss level affect the relative perceptibility of this artifact. The efficacy of the model developed in this research suggests that certain factors regarding image content impact the relative perceptibility of gloss differences in pictorial imagery. Not surprisingly, the length of the edges between high and low gloss areas and the relative straightness of these edges were major factors determining the perceptibility of differential gloss. It is also not terribly surprising that images with centrally located differences in gloss ranked higher for perceived differential gloss than those with differences closer to the periphery. For this reason, single image features that are large enough to reach into the periphery have less impact than those contained entirely in the center of the image, indicating an effect of size. It was also true that features that were smaller than about 0.5 cm had less impact than those up to about 10 cm. What is less intuitive, perhaps, was that a single feature, in the proper size range, had more impact than lots of features. Moreover, what may have been most surprising was that the presence of human faces affected the perceptibility of differential gloss in this experimentation. It is possible that the attention naturally given to faces caused observers to inspect these areas and areas near the faces having high differential gloss more closely.

ACKNOWLEDGMENTS

The author wishes to thank Chunghui Kuo for his help in the initial experiment in this study, the two reviewers who supplied insightful and valuable comments that led to a stronger paper, and all of the observers who participated in the testing.

REFERENCES

[1] International Committee for Information Technology Standards, "What is INCITS?", http://www.incits.org (2004).

[2] S. P. Farnand and D. R. Nilosek, "Measurement of differential gloss using a μ-goniophotometer", *Proc. IS&T's NIP23: International Conference on Digital Printing Technologies* (IS&T, Springfield, VA, 2007), pp. 368–372.

[3] Y. S. Ng, C. Cui, C. Kuo, E. Maggard, P. Morris, and D. Mashtare, "Gloss uniformity measurement update for ISO/IEC 19751", Proc. SPIE **5668**, 163 (2005).

[4] Y. S. Ng, E. Zeise, D. Mashtare, J. Kessler, J. Wang, C. Kuo, E. Maggard, and P. Mehta, "Standardization of perceptual based gloss and gloss uniformity for printing systems (INCITS W1.1)", *Proc. IS&T's PICS Conference* (IS&T, Springfield, VA, 2003), pp. 88–93.

[5] R. S. Hunter and R. W. Harold, *The Measurement of Appearance*, 2nd ed. (Wiley, New York, 1987).

[6] J. S. Arney and D. R. Nilosek, "Analysis of print gloss with a calibrated micro-goniophotometer", J. Imaging Sci. Technol. **51**, 509–513 (2007).

[7] G. Obein, K. Knoblauch, and F. Viéot, "Difference scaling of gloss: Nonlinearity, binocularity, and constancy", J. Vision **4**, 711–720 (2004).

[8] W. Ji, M. R. Pointer, M. R. Luo, and J. Dakin, "Gloss as an aspect of the measurement of appearance", J. Opt. Soc. Am. A **23**, 22–33 (2006).

[9] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg, "Toward a psychophysically-based light reflection model for image synthesis", Comput. Graph. **34**, 55–64 (2000).

[10] R. O. Dror, A. S. Willsky, and E. H. Adelson, "Statistical characterization of real-world illumination", J. Vision **4**, 821–837 (2004).

[11] ISO 20462-2:2005/Cor 1:2007, http://www.iso.org

[12] L. L. Thurstone, "A law of comparative judgment", Psychol. Rev. **101**, 266–270 (1994) (original 1927).

[13] C. J. Bartleson, "Thresholds and matching", in *Optical Radiation Measurements*, edited by C. James Bartleson and Franc Grum (Academic, New York, 1984), Vol. **5**, Chap. 7, pp. 367–440.

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

[15] G. T. Buswell, *How People Look at Pictures* (University of Chicago Press, Chicago, 1935).