A STUDY OF VISUAL PLEASINGNESS AND COLOR MATCH OF PICTORIAL IMAGES

Angelica Li
angelica.c.li@gmail.com

KEYWORDS

colorimetry, image quality, preference, subjective

ABSTRACT

The perception of visual pleasingness is a subjective response based on the merit of an image alone, while color match is a subjective response based on the comparison between an image and a reference. To study the difference between the subjective choice of pleasingness and color match and if quantitative parameters may serve as useful predictors, three pictorial images were used as input materials and output to four calibrated printing devices: sheet-fed offset, drop-on-demand inkjet, continuous inkjet, and electrophotographic. The two workflows used were: (1) to print to the full gamut of the device to achieve visual pleasingness, and (2) to print to match the reference device to achieve color match via profile conversion.

This paper details research in which seven observers performed two paired comparison tests to (1) select the more pleasing image and (2) select the closest matching image to a provided reference. In terms of visual pleasingness, the results show that the sheet-fed offset image was judged as the most pleasing image while the ranking of the other devices varied. For closest match to a reference, an offset image was judged visually as the best visual color match to the provided reference. Results also indicate that quantitative parameters (e.g., chroma, hue, paper brightness, color gamut) do not correlate to visual pleasingness of pictorial images. However, there is a correlation between addressability and visual pleasingness. Findings also indicate that color difference (ΔE) between colors of interest in the reference and the sample can be a useful indicator of color match to a reference image.

1. INTRODUCTION

The visual evaluation of prints is a subjective process, which differs depending on the purpose of the evaluation. When print buyers are judging color proofs, the criterion is based on visual pleasingness, which is highly influenced by an individual's preference or bias. When print buyers perform press-side color approval, the criterion is based on color match between the press sheet and the color proof.

The criteria used to evaluate prints for visual pleasingness and color match are both subjective in nature. The question is, "Can the criteria used for visual evaluation be correlated with quantitative device- and image-based parameters?"

2. LITERATURE REVIEW

The visual evaluation of pictorial images and color reproduction is heavily shaped by subjective influences and is not backed by strong scientific findings (Field, 2004, p. 318). In color approval, Field states that there is "no one objective goal that will produce the 'correct' result" (p. 323), therefore "variability… is the norm." This statement applies to all visual evaluation and describes how it is difficult to obtain predictable results from a subjective process. Observers form opinions that are more emotional than rational (p. 318) and there is a necessity for a more structured and scientific approach to the formation of these opinions (Hunt, 2004, p. 163). Paired comparison testing is one such approach, allowing for the quantitative assessment of the subjective difference between two images as perceived by an observer or judge (p. 163).

The numerous subjective influences on visual evaluation include those based on the observer and the specific situation. The physiological and psychological characteristics of the observer have an effect on visual evaluation results (Field, 1998, p. 131). For instance, trained and untrained observers may place more or less emphasis on different areas in an image (Hunt, 2004, p. 163) and also the individual's perception of "excellence" will shape what they judge as a good image (Field, 1998, p. 131). The picture content of the image, its end use, and the purpose of the evaluation also affect observer perceptions (p. 131) — observers will respond differently if they are looking for a pleasing image versus one that best matches a reference.

"Pleasingness" is defined ambiguously as something that gives pleasure, or something agreeable and liked by the senses; the visual pleasingness of a printed image is not clearly defined and may be interpreted differently depending on the observer. There is less room for interpretation when determining if an image matches a reference because it requires an observer to select an image based on its similarity or exactness to the provided reference. Studying how observers respond visually when looking for pleasingness or color match is a crucial part of understanding the various factors that come into play when observers evaluate pictorial images.

Previous studies have revealed patterns in judgment variability and preferred color reproduction. Research in black and white tone reproduction found numerous variables in individual perception and judgment. For instance, there is variation when the same observer judges a reproduction at different times. Another cause of variability, or lack of agreement amongst judges, is an observer's bias for or against the subject matter in the image. (Field, 2004, p. 319)

In terms of preferred color reproduction, examples in research has shown how color preference for certain objects in reproduced images may differ from real-life color. Hunt describes a preference test for the quality of color reproduction in reflection prints that indicated that the preferred skin color and grass color is more yellow (sun-tanned, and brighter) than the average real skin or grass color. For blue sky, the preferred hue is similar to its real life color; however, the preferred color has a much higher purity (greater chroma) (p. 176). Considering the numerous influences on visual evaluation, these observed preferences should be regarded as examples (p. 177), however, generalized preferences could be gleaned through further study.

3. OBJECTIVES

This research seeks to provide insight into the following three research questions:

1. Will printing technology and image content contribute to the subjective evaluation of pictorial images in terms of visual pleasingness?

2. Will printing technology and image content contribute to the subjective evaluation of pictorial images in terms of color match to a reference?
3. Do device-based or image-based quantitative parameters correlate with subjective rankings?

4. METHODOLOGY

The following section details the procedures used to investigate the three objectives of this research.

4.1 PAIRED COMPARISON TESTING

Paired comparisons were used to rank three sets of pictorial color images produced by different printing technologies. The test for Objective 1 was on the basis of visual pleasingness; the test for Objective 2 was on color match to a reference.

Paired comparison tests were completed in a GTI viewing booth under standard D50 lighting and recorded in a customized spreadsheet that handles statistical analysis using four samples. Samples were mounted onto neutral gray board. To eliminate the factor of gloss from the subjective evaluation, tests were performed with a clear plastic sheet over all samples (Figure 1). To prevent external influences on color judgments and reduce light contamination, all ambient lights were off. As strongly colored objects may "distort" the color temperature of viewing conditions (Field, 2004, p. 318), observers wearing strongly colored clothing were asked to cover themselves with a neutral shirt.

Figure 1. An example of paired comparison setup for visual pleasingness.

Observers were sampled from university-level students in graphic art related fields ranging from print science, to photography, and new media arts. The majority of observers (five out of seven) had no previous image evaluation experience, while two observers had briefly worked in the printing industry in a lithographic pressroom setting. The sampling included four male observers and three female with no known color deficiencies (deficiency tests were not performed, however observers stated that they passed the Ishihara Test for Color Blindness on previous occasions.)

The procedure for the test was: observers were approached and asked if they could spare 15 to 20 minutes for a visual image evaluation experiment. The purpose of the test was not discussed until after judgments were completed. For the two paired comparison tests, four samples were labeled A through D for each of the three images, resulting in 12 total samples. During the test, each sample was compared to the others in randomly sequenced pairs (A-B, A-C, A-D, B-C, B-D, C-D) for a total of 18 side-by-side comparisons.

In the test for visual pleasingness (Objective 1), the observer was asked for each pair: "Which of the two prints do you find most pleasing in terms of color?"

In the test for color match (Objective 2) a sheet-fed printed sample was provided as the reference. The observer was asked for each pair: "Which of the two prints is a closer match to the reference in terms of color?"

After judgments were complete, observers were asked to comment on how they made their observations and what portions of each image they focused on the most. These comments were used to select the key color swatches for quantitative analysis.

4.2 PRINTING TECHNOLOGY

The IT8.7/4 characterization target was used to create printer profiles for output in this research. Targets were printed without applying curves, and measured using an X-Rite Eye-One iSis spectrophotometer and X-Rite/Gretag MeasureTool software. Printer profiles were generated from spectral measurements in X-Rite/Gretag ProfileMaker 5.0 software.

Four devices were used to output the sample prints in this research:

- Sheetfed offset lithographic (SF)
- Electrophotographic (EP)
- Inkjet (IJ_1)
- Inkjet (IJ_2)

All prints were on similar coated text stocks. The two workflows were: (Objective 1) printing to the device's full gamut, and (Objective 2) printing color-managed images to match the reference device via profile conversion.

4.3 PICTORIAL IMAGES

The Adobe RGB test images printed for visual evaluation were the knife, bread, and produce images from the PCRI 2 (Pictorial Color Reference Images, second series) image-set (Figure 2). These images were selected due to the presence of memory colors, for instance, steel grey and common food colors such as tomato red, cheddar orange, and lettuce green. These colors make these images good for visual assessment because people have established pre-conceptions on what memory colors should look like, and errors (such as in hue) are more serious and noticeable (Hunt, 2004, p. 167).

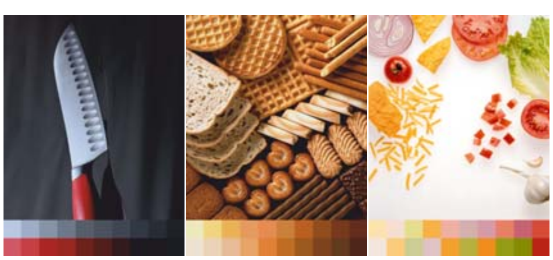

Figure 2. PCRI 2 images used in this lab (from Left: Knife, Bread, and Produce).

Twenty-four printed images were collected for this research, and included four samples (SF, EP, IJ_1, IJ_2) of three images (knife, bread, and produce) for each objective. During paired comparison tests, color patches on printed images were covered giving a viewed image size of 3 inches wide by 3.4 inches high.

4.4 DEVICE- AND IMAGE-BASED COLORIMETRIC ANALYSES

To approach Objective 3, device-based and image-based colorimetric parameters will be analyzed (including hue, chroma, color gamut, ΔE, and addressability). Observations will be made to determine if quantitative parameters can be indicative of visual pleasingness or selection of color match. A limitation to this is that there must be a strong agreement among the observers if quantitative parameters are to correlate to subjective visual evaluation.

5. RESULTS AND DISCUSSION

The following section outlines the results of the paired comparison test for visual pleasingness, measured hue and chroma for key color swatches, and other criteria for pleasingness.

5.1 OBJECTIVE 1: VISUAL PLEASINGNESS

The paired comparison utilized four samples of each image. Seven judges were asked: "Which of the two prints do you find most pleasing in terms of color?"

5.1.A PREFERRED PRINTING TECHNOLOGY

For paired comparison results, judges who are consistent have no triads, or no conflicting comparison judgements. W is the coefficient of concordance, which is the agreement between judges (0= no agreement, 1= perfect agreement). Real difference indicates that a print has been judged as containing differences from the other samples.

All seven judges were consistent in the test for visual pleasingness. For the knife image, there was minimal agreement (W=0.31) between judges, and a real difference in the SF print. For the bread image, there was no agreement (W=0) between judges and no real differences in any print. For the produce image, there was minimal agreement (W=0.25) between judges and no real differences in any print.

The sheet-fed sample was ranked as the most preferred; there was no clear pattern in preference for the other devices (low agreement or large variability amongst judges). The ranking for each print is shown in Table 1.

Table 1. Ranking for visual pleasingness.

	1st	2nd	3rd	4th
Knife	SF	EP	IJ_2	IJ_1
Bread	SF	IJ_1	EP	IJ_2
Produce	SF	EP	IJ_1	IJ_2

The paired comparison test revealed that sheet-fed offset lithographic printing technology produced the most visually pleasing images. To determine what quantitative parameters influence visual pleasingness, device- and image-based factors were analyzed in all samples.

5.1.B DEVICE-BASED ANALYSES

Device-based quantitative parameters include paper brightness, device color gamut, and addressability (Table 2). The most preferred print (SF technology) had the smallest gamut and was not on the brightest paper, indicating that the these two parameters do not correlate to visual pleasingness. The exception is that there is an observed positive correlation between addressability and visual pleasingness.

5.1.C IMAGE-BASED ANALYSES

Hue and chroma are image-based parameters. Key color swatches were used to measure the hue and chroma of the printed samples (Tables 3 and 4). These colors were chosen based on judges' comment on what areas they focused on the most.

In the image-based colorimetric analysis of hue and chroma, there were no strong patterns, indicating that there is no correlation between hue

Table 2. Device-based quantitative parameters.

Quantitative Parameters				
	SF	EP	IJ_1	IJ_2
Paper Brightness	86%	88%	87%	83%
Gamut Volume	384,251	509,099	457,730	561,664
Addressability (dpi)	2400 x 2400	600 x 600	600 x 600	1440 x 720

Tables 3 and 4. Image-based quantitative parameters (hue and chroma).

Hue (h)				
Color	SF	EP	IJ_2	IJ_1
Knife Background	268	276	212	244
Knife Blade	271	299	210	138
Color	SF	IJ_1	EP	IJ_2
Bread Waffle	75	73	71	78
Bread Slices	69	82	72	83
Color	SF	EP	IJ_1	IJ_2
Produce Lettuce	102	99	104	97
Produce Tomato	34	36	42	42

h of most preferred print

Chroma (C*)				
Color	SF	EP	IJ_2	IJ_1
Knife Background	7	8	5	3
Knife Blade	12	12	9	2
Color	SF	IJ_1	EP	IJ_2
Bread Waffle	51	68	55	65
Bread Slices	10	39	11	17
Color	SF	EP	IJ_1	IJ_2
Produce Lettuce	49	51	49	56
Produce Tomato	63	66	63	67

Lowest C* | Highest C*

angle or chroma and visual pleasingness. This is because visual pleasingness is highly subjective depending on the individual. The "correct" rendering of memory colors is weighted heavily in pleasing preference, however there is no consistent correct hue. For instance, for the bread image, half of the judges chose warm (toasted) bread as the most pleasing, while the other half preferred cooler (non-toasted) bread. This clearly illustrates that individual bias towards image content will affect judge preference.

This research focuses on quantitative (instrument-based) parameters that approximate subjective responses. For instance, if it is observed that tomato red is most pleasing at a hue angle of 34° and chroma of 63 then those numbers can serve as an aim point for controlling print production.

Paired comparison techniques, while useful in depicting human visual responses, are limited in its ability to correlate to specific causes. Also, if there is little to no agreement between judges, then rankings are merely arbitrary. Therefore, other psychometric analysis techniques should be explored.

5.2 OBJECTIVE 2: COLOR MATCH

The paired comparison utilized four samples of each image and a sheet-fed printed sample was provided as the reference. Seven judges were asked: "Which of the two prints is a closer match to the reference in terms of color?"

5.1.A PRINTING TECHNOLOGY MATCH

All seven judges were consistent in the test for color match to a reference and the IJ_1 print contained real differences for all three images. Between judges, there was average agreement (W=0.48) for the knife image, okay agreement (W=.62) for the bread image, and good agreement (W=0.71) for the produce image.

There was a clear consensus (average to good agreement between judges) in ranking for best match to a sheet-fed reference (Table 5).

Table 5. Ranking for color match to a reference.

	1st	2nd	3rd	4th
All Images	SF	IJ_2	EP	IJ_1

The paired comparison test revealed that the SF print was the best match to the provided sheet-fed reference. This is as expected because the SF sample and the reference were printed in the same press run, therefore the color match between them is invariant.

Although there is an average to good agreement between judges in ranking, the visually determined color match can vary and is highly dependant on the color management applied to match the reference.

5.1.B DEVICE-BASED ANALYSES

Device gamut is important for color matching because the color gamut of the sample printing devices must be large enough to encompass the gamut of the reference to be matched. In this research, the images used were well within all gamuts; therefore, there was no significant influence on matching choice based on device gamut size.

Also, paper brightness had no significant effect on the selection of a match. The second ranked match to the reference (IJ_2) had the lowest brightness (83%), with minimal OBA content.

Addressability influences the choice for best match (though not necessarily color match) as the sharpness and effective resolution of the device affects the distinct edges and fine details of an image. Therefore, addressability that closely matches the reference is preferred even when looking solely at color. (There are also other variables to consider when looking at addressability beyond mere dpi counts, such as dot integrity and gray levels). The second ranked match to the reference (IJ_2) was printed on a device which had a dpi of 1440 x 720, while the reference and SF sample had a dpi of 2440 x 2440.

5.1.C IMAGE-BASED ANALYSES

Key color swatches were used to evaluate the color difference (ΔE_{ab}) between the printed samples and the sheet-fed reference (Table 6). These colors were chosen based on judges' comment on what areas they focused on the most.

Table 6. Color difference for key color swatch samples compared to the reference.

Color	SF (ΔE)	IJ_2 (ΔE)	EP (ΔE)	IJ_1 (ΔE)
Knife Background	0.0	6.3	4.2	22.1
Knife Blade	0.0	2.6	2.7	16.8
Bread Waffle	0.0	5.5	2.6	5.4
Bread Slices	0.0	4.7	1.9	8.6
Produce Lettuce	0.0	3.9	8.9	11.1
Produce Tomato	0.0	5.6	3.8	4.5

Excellent Match (0–2 ΔE)	Fair Match (6–10 ΔE)
Good Match (2–6 ΔE)	Poor Match (>10 ΔE)

Looking at the key color samples, color difference played a significant role in the selection of best match to the provided reference. The first ranked match (SF) had the smallest color difference (0 ΔE) and the second ranked match (IJ_2) had good or acceptable color matches (between 2–8 ΔE). Both the IJ_2 and EP samples had mostly good matches, but the EP sample had slightly smaller color differences. The EP sample was ranked third behind the IJ_2 sample, indicating that there may be other parameters affecting the choice for best color match along with ΔE measurements.

In terms of image-based colorimetric analysis, color differnce (ΔE) can be a useful indicator of color match. A more complete analysis (more color patches) of the colors in an image will provide a better approximation.

Though contrast and uniformity were not measure in this research, it was observed that they are influential when color is similar between two prints. The matching of overall contrast and image quality was prioritized for match over perfect color. Notice that color difference was at times slightly higher in the selected second best match (IJ_2). However, it may have been chosen as the best match because the overall contrast, uniformity, and image quality was a closer visual match to the sheet-fed reference, affecting the judges evaluation of color match.

6. CONCLUSION

This research examines how image quality criterion, printing technology, and image content can impact subjective image quality in terms of visual pleasingness and color match of an image. It also attempts to identify quantitative colorimetric parameters as useful predictors of subjective visual responses. In this research there is an observed positive correlation between addressability and visual pleasingness— higher addressability is preferred. However, despite the correlation in addressability, findings in general indicate that there are no comprehensive quantitative parameters that can successfully predict the visual pleasingness of pictorial images. There is also an observed negative correlation between color difference (ΔE) and visual color match— as color difference increases, color match decreases.

7. FUTURE RESEARCH

Observations between visual preference and objective measurements were not intended to describe direct correlations between the two. Rather, they were used to identify whatever patterns appeared within the small sampling sizes of this research.

This research provides very general patterns in factors that may influence an observer's visual response for pleasingness and color match. It is important to understand that the correlation of subjective responses to quantitative data is a tricky subject. This is especially true for color because it is three-dimensional (Hunt, 2004, p. 636) and is an extremely complex and subjective visual stimulus. Therefore, if there is any chance for patterns to be more significant or to provide generalizations for an average observer, future research must utilize a large quantity of images of varying content, and testing must involve a sufficiently large sample size from a diverse population (taking environmental and demographic factors into account). Patterns and generalizations in visual color match are easier to identify as there is less room for subjective sway.

Other psychometric evaluation techniques should be used to provide more differentiating power in visual evaluation (whether it be for pleasingness or color match). For instance Farnand's research in image quality evaluation looked at the perceived value of prints using dollar values as an indicator for visual responses to image quality parameters (Farnand, 2008).

In addition, further evaluation of color within images is needed for a more complete analysis. This research only looked at a few key color patches which is not sufficient for providing significant conclusions. By collecting more colors of interest within the image, e.g., a PCRI chart, we can analyze the ΔE distribution as a cumulative relative frequency (CRF) chart to approximate color match in images.

8. ACKNOWLEDGMENTS

The author would like to thank Professor Robert Chung for his dedication to the subject of this study, and for providing direction and a clear vision. Many thanks also go to Professor Franz Sigg and Professor Edline Chun for their editorial guidance. Last, but not least, the author would like to thank her fellow Print Media peers who have provided endless technical and moral support.

9. REFERENCES

Farnand, S. (2008). *Minding the Gap: Evaluating the Image Quality of Digital Print Technologies Relative to Traditional Offset Lithography* (Monograph No. PICRM-2008-08). Rochester, NY: Printing Industry Center at RIT.

Field, G. (1998). Color Approval in the Graphic Arts. In R. Buckley (Ed.), *Recent Progress in Color Management and Communications* (p. 127–132). Springfield, VA: Society for Imaging Science and Technology (IS&T).

Field, G. (2004). Color Communication. In *Color and Its Reproduction* (3rd ed., p. 320–325). Pittsburgh: GATFPress.

Hunt, R. (2004). *The Reproduction of Colour* (6th ed.). Chichester, England: John Wiley & Sons, Ltd.