CrowdColor: Crowdsourcing Color Perceptions Using Mobile Devices

Jaejeung Kim[1], Sergey Leksikov[1], Punyotai Thamjamrassri[2], Uichin Lee[1], Hyeon-Jeong Suk[2]
[1]Department of Knowledge Service Engineering, KAIST
[2]Department of Industrial Design, KAIST
{jjk, sergey, uclee, punyotai, h.j.suk}@kaist.ac.kr

ABSTRACT
Providing accurate color information to online shopping customers is important for their purchase decisions. However, due to the multiple imaging processes that product photos undergo, end-users often experience a color mismatch between the color of the photo online and the product received. Therefore, we use a crowdsourcing approach to generate what we term *CrowdColor*, which is the collective color reported by individuals using a mobile color picker. CrowdColor serves as a color review application from the customers' perspectives in the form of a color palette that represents the product color. We perform controlled experiments to evaluate the accuracy of CrowdColor and to understand how the effects of the device and lighting conditions may influence the crowd's color perception and input tasks. The quantitative results reveal that CrowdColor achieves high accuracy and is positively rated overall. Based on experimental analyses, we present design guidelines for crowdsourcing color perception tasks.

Author Keywords
Crowdsourcing; color perception; graphical perception; mobile shopping

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI):

INTRODUCTION
A cultural shift toward a mobile-centric lifestyle has expanded the mobile shopping market. Smartphones and tablet devices now serve as a gateway for customers to search for and purchase products online. Most recent online shopping malls fully support mobile pages that enable customers to browse, compare, and purchase various products. Despite the convenience of online shopping, its virtual characteristics increase the level of difficulty, preventing some customers from finding and understanding detailed information about products. In particular, regarding fashion items (e.g., T-shirts, pants, earrings), many customers cannot understand the "real color" information from the product image online due to distorted colors in images. While color is a primary purchase criterion in the area of fashion, inaccurate color representations in e-commerce can lead to negative shopping experiences, such as complaints and product returns [6].

Thus, providing accurate color information is crucial in the online shopping experience. Most major online shopping malls provide customer review opportunities, and such reviews are a key source of subjective opinions about products. However, unlike other characteristics such as size and quality, which can be described with simple words (e.g., big/small, high/low), it is not easy to convey the precise meaning of color using text. Customer reviews that include a photo are also prone to color distortion, as described above, during the image-generation process.

In order to address the complex problem of color mismatches between the color of a photo online and the actual product, we develop an approach which crowdsources color information by receiving explicit color input from customers. We aggregate the individual instances of color input from the crowd to generate a collective color that represents the product, which we refer to as *CrowdColor*. In this paper, we experimentally assess the feasibility of generating CrowdColor. Because color perception is predominantly affected by environmental factors (i.e., lighting) and display characteristics, we conducted a controlled experiment using a 4×2×2 factorial design. We chose four representative colors (red, blue, yellow, and gray) with participants who inputted the perceived color under two different light settings using two different display device types in order to observe the effects of these variables.

We begin with reviewing related work and presenting a motivating scenario for CrowdColor. We then describe the design of CrowdColor in detail. Next, we evaluate CrowdColor and observe how the environmental factors affect the accuracy of the color. Finally, we discuss the feasibility of CrowdColor and derive system design guidelines for crowdsourcing color perception tasks.

MobileHCI '15, August 24 - 27, 2015, Copenhagen, Denmark

DOI: http://dx.doi.org/10.1145/2785830.2785887

RELATED WORK

Color Reproduction

Color fidelity is endangered by multiple factors, ranging from image capturing to delivery. An imaging engine in a camera creates a digital image for capturing, which is largely influenced by the lighting sources (e.g., daylight and florescent light). For delivery, an image rendering engine processes an image and attempts to provide the best visualization on a given display type.

Numerous approaches have been proposed to correct and compensate for color distortion and to obtain higher color fidelity. *Color Match* by Jain *et al.* tackles the problem of distorted image colors, either from inappropriate lighting conditions or from poor imaging from the beginning [5]. In their cosmetics scenarios, a user's original skin color is inferred with color compensation, requiring a self-photograph by the user while they hold a color chart. However, this requires the user to use a color chart at the time of capture, which is not feasible for existing photos that were captured without this chart.

Other approaches suggest solutions for improving poor image rendering due to the display type and/or environmental characteristics [7, 8]. In particular, the display type and lighting conditions were considered in an application of adaptive tone mapping. However, this method is applicable only when the original image is correctly captured.

Previous approaches require heavy image processing and external tools to produce a desired original color. However, most online stores cannot afford to apply these advanced methods to their webpages. Furthermore, these methods cannot correct a photo if it was intentionally color-edited, e.g., to make the product look more appealing.

Crowdsourcing Graphical Perceptions

Crowdsourcing is a cost-effective means of performing online tasks. Heer *et al.* made an assessment of Amazon's Mechanical Turk when performing the graphical perception tasks of spatial encoding and luminance and contrast judgment [3]. They limited their work to simple graphical judgment tasks. Color perception was neglected owing to the lack of control over the display settings.

Meanwhile Lin and Hanrahan harnessed the crowd to extract color themes from images [2]. However, their crowd workers were given a limited number of color candidates to choose from. They used 40 color candidates that were automatically extracted out of an image. Also, their color comparison and selection task did not involve any physical object and was performed solely on the same display.

No study has attempted to use crowd workers to extract a color from a physical object and translate it into a digital color. Such a task is challenging due to environmental and device limitations. However, we carefully approach this challenge in the later sections.

MOTIVATING SCENARIO

Sarah often browses online shopping malls for clothes. Once she finds a skirt she likes, she searches for the same skirt in different shopping malls in order to compare the look and color. She finds that the color is inconsistent across the malls; even the colors in a series of images posted by another seller differed. Sarah becomes confused as to which photo shows the "true color." She scrolls down to the customer reviews to see what other customers who bought the same item wrote about the color, where she finds entries such as "the color is darker than the photo," and "it's more of a blueish gray." Sarah imagines what the real color would be like, but remains unsure of the exact color. Some customer reviews had photos attached; the colors in those photos varied as well.

Next, Sarah visited a shopping mall that supported CrowdColor. She was able to see the color palette selection made by customers who purchased the product. Sarah immediately understood the skirt color without needing to infer it. This shopping mall appeared more trustworthy because she knows that some shopping malls edit the product images in order to make them more appealing. Furthermore, Sarah knows that the CrowdColor entries are from customers who have used the application after actually purchasing the item. Thus, Sarah purchases the skirt, finding that the actual color matched the CrowdColor results that she saw on her mobile display. Sarah also left a CrowdColor review for future customers.

CROWDCOLOR SYSTEM

Color Input

CrowdColor is a color selection and representation application in the form of a color palette. It is designed to receive color inputs from individuals' color perceptions of a real object, and it aggregates these perceptions into one collective color. In order to receive color inputs from the crowd, we designed a color selection interface with which mobile users can easily report a color. We implemented a mobile web service that manages the color inputs from multiple users on mobile web browsers.

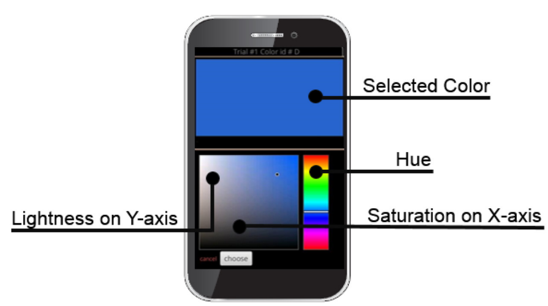

Figure 1. Color picker (bottom half) and selected color (top)

The user interface is divided into two sections: a lower screen and an upper screen section (Figure 1). The lower half of the screen is the Adobe Photoshop color picker, which was selected because its control for hue was independent of the saturation and lightness controls. Hue is considered to be more important because it is the easiest and primary characteristic to which people react. Independent control of the hue reduces the dimensions of the color selection process, which simplifies the color selection as intended. While navigating the colors in the lower half, the selected color can be clearly seen on the upper half. The color selection can be made using one hand, leaving the user free to hold the object in the other hand and compare the color.

CrowdColor Generation

The user selects the matching color and this is sent to the system: the system receives the RGB data and aggregates the RGB inputs into the corresponding object ID. Then, the RGB value is aggregated with previous users' input instances for the same item (color) by obtaining the average of the previous inputs.

Light and Device Adaptation

According to the color science literature, color presentation and perception are significantly influenced by environmental, device, and human factors. According to an earlier study [9], the following factors are primarily responsible for color perception differences.

- Lighting: The color of the product will differ under different light temperatures; e.g., daylight, florescent light, and incandescent light result in different color perceptions.
- Device: Different device displays have different color gamuts. These are a subset of colors that can be accurately represented on a display, and display hardware differences cause these differences. For example, on mobile devices, two types of flat panels, IPS and AMOLED, are widely used. These two types have different gamut ranges.
- Human: Chromatic adaptation is a phenomenon which induces a color perception illusion after the adjustment of the eyes on a certain color, making another color appear differently from how it should.

Because human factors are uncontrollable within the proposed CrowdColor system, we limit the scope of interest here solely to the lighting and the device. Therefore, the color input for the same object ID varies according to the display device type and lighting conditions at the time of their input. Current mobile phones are capable of detecting device types using simple JavaScript. Lighting conditions are also sensible using the camera and illumination sensor. Through clustering, inputs made from the same device with similar lighting conditions can be aggregated for higher accuracy. However, we did not fully implement this feature; instead, we conducted a controlled in-lab experiment in order to demonstrate the feasibility of this feature.

EXPERIMENT 1: CROWDSOURCING COLOR INPUTS

The first experiment was conducted in order to collect the input sources required to generate the CrowdColor output. We also assessed the accuracy, which in this case denotes the crowd worker's color perception and selection ability to translate the object color into a digital color.

We used standardized color papers as stimuli instead of an actual object (e.g., a T-shirt) in order to achieve more objective measurements. We specifically targeted the most influential factors that contribute to color perceptions, i.e., the lighting condition and display device type. Furthermore, by crowdsourcing the color inputs from the participants, we were able to generate the CrowdColor. The color accuracy of the CrowdColor is evaluated in the following section.

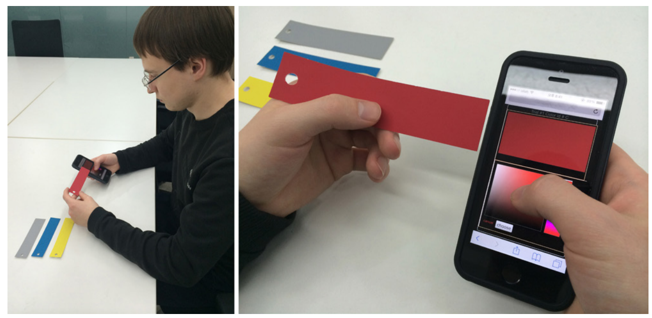

Figure 2. Color input task using a mobile device; four color stimuli and two mobile devices were given

Settings

We conducted a 4 (colors) × 2 (lighting conditions) × 2 (devices) factorial design experiment. In order to examine and measure the effect of each variable carefully, we conducted the study in a lab setting.

For the color stimuli, we chose red, yellow, and blue from the primary colors and gray from the neutral colors. These are basic colors that we encounter in our everyday lives. The factorial experiment design prevented us from using more color stimuli.

We drew these four colors from a color atlas paper (accredited by the Korean industry standard). The color atlas consists of 155 colors and is based on the hue and tone system, which is commonly used by color scientists. The lighting conditions were designed to match daylight (5400K) and warm florescent light (3800K) in order to encompass both daytime and nighttime settings. The brightness ranged from 310 to 400 lux, whereas typical household brightness levels range from 300 to 700 lux. For the display devices, we selected the two most common mobile devices available on the market: the iPhone 5S and the Galaxy S4. The iPhone uses an IPS display panel, whereas the Galaxy uses an AMOLED panel; this also provides a good comparison between display panels and different gamuts. IPS and AMOLED devices differ in their color representation mechanisms. Most smartphones use one of these types.

Method

We recruited 31 people (19 males, 12 females), whose ages ranged from 19 to 30 years (mean = 24.4). The participants were provided with US$10 compensation for their participation. Before the experiment, the participants were tested for color vision deficiency (CVD) using the Farnsworth-Munsell 100 test [1]; the finding showed that they were within the range of no CVD to slight CVD, which is within the natural demographic distribution.

Each participant was given four color stimuli and two devices and was instructed to choose the corresponding color on the device (Figure 2). In order to prevent a learning effect from repeated inputs of the same color sequence, we counter-balanced the order of the color stimuli. Before this experiment, a five-minute trial session was undertaken in order to allow the participants to adapt to the environment and to the color picker. The participants were then asked to select the color closest to that of the color stimuli on the mobile color picker. The participants performed four input trials using each color stimuli. Each participant underwent eight trials using each device. After each set of four inputs, the lighting condition was changed. We also recorded the participants' completion time per input for analysis.

We measured each individual RGB input using a spectrophotometer (Konica Minolta CS-100A). A spectrophotometer is a color detection device which measures an illuminating or reflecting object's wavelength. We attempted to measure the color the participant would have perceived by holding the spectrophotometer at the level of the eyes and operating the spectrophotometer toward the target color.

The CIELAB color is a perceptually uniform color space; thus, it was used for quantifying the perceptual color differences between the colors [4]. Based on CIELAB data transformed from the RGB values, we acquired the color distances (ΔE) between the color stimuli and the color inputs using a CIEDE2000 formula. Here, ΔE denotes the Euclidean distance between two colors with a subset of value compensation.

Results

CrowdColor Generation

In total, 496 trials were conducted. From these inputs, we generated two types of CrowdColor. The first was the device and light-adaptive CrowdColor (Adaptive CrowdColor: ACC), which only aggregated the inputs with the same device and lighting conditions (Table 1). The second was the opposite: reverse mapping of the CrowdColor (RCC), where the user input on the iPhone under florescent light was rendered on the Galaxy viewed under daylight (Table 1). In this manner, we created the best performing case and the worst performing case in order to compare the accuracy gap.

All color samples in Table 1 were cropped from one image in which the color stimuli and all CrowdColors rendered on the two devices were placed together. This way, we could visually compare the relative difference and gain a rough idea of the color similarities/differences. However, note that this is yet another image captured by a camera which distorted the "real color." The color distance can only calculated by a spectrophotometer. Here, the measured ΔE is shown on the right.

The best performing case was an ACC case that exhibited accuracy as high as $\Delta E = 2.00$ for red (B1). This is less than the JND (just noticeable difference) with $\Delta E \approx 2.3$, which is the threshold that is scarcely distinguishable for the human eye. However, for gray, it was as low as $\Delta E = 20.00$ (C4). This is clearly noticeable and would appear significantly different. However, as shown in Table 1, most colors in the ACC type were visually very similar to the stimuli.

We compared the ACC and RCC types in terms of the color distance. As shown in Table 2, all conditions outperformed RCC, except for one. However, the difference was not as high as expected.

Table 1. Generated CrowdColors (A~D) compared with four color stimuli (S1~4) for color accuracy measurements under the 5400K lighting condition

Stimuli	Type	Galaxy S4 / iPhone 5s	Color Distance (ΔE)
S1	ACC	A1	3.7
		B1	2.0
	RCC	C1	6.7
		D1	5.1
S2	ACC	A2	8.0
		B2	17.5
	RCC	C2	10.1
		D2	19.5
S3	ACC	A3	9.7
		B3	7.8
	RCC	C3	14.7
		D3	12.1
S4	ACC	A4	14.4
		B4	13.7
	RCC	C4	20.0
		D4	15.7

Effect of Color, Light, and Device on Color Accuracy

In order to observe the effects of the color stimuli, light, and device, which influence the color accuracy, we conducted a three-way ANOVA. First, both color ($F_{3,473} = 53.25$, $p < 0.001$) and light ($F_{1,473} = 36.33$, $p < 0.001$) had a significant effect on the color accuracy, but not on the device ($F_{1,473} = 2.66$, $p = 0.10$). This primarily results from the color perception tolerance level for different colors. A vivid color (e.g., red) was more accurately crowdsourced relative to yellow and blue. Furthermore, the two lighting conditions significantly changed how people perceived the colors. However, the device did not independently affect the color accuracy, which indicates that regardless of the device, the participants were able to generate similar colors (despite the fact that the RGB values differed due to device differences).

There was a significant interaction effect of the color and light on the color accuracy ($F_{3,473} = 3.75$, $p < 0.05$). This explains that the lighting conditions changed the color stimuli and that color already exhibited a significant effect on the color distance. However, the interaction effect of the color and device on the color accuracy ($F_{3,473} = 2.63$, $p = 0.05$) was marginally significant. The two display types (AMOLED and IPS) have different color gamuts and brightness levels, but consistent behaviors were not observed. The accuracy was more dependent on the color itself. This result is in agreement with other ΔE analysis results, where the iPhone was superior to the Galaxy when selecting red, whereas it was inferior for yellow.

Table 2. Color difference between adaptively mapped and reverse mapped CrowdColor for each color stimulus

	Galaxy		*iPhone*	
	3800K	**5400K**	**3800K**	**5400K**
Red	1.97	2.97	6.65	3.06
Yellow	3.14	2.07	11.06	2.03
Blue	5.61	4.99	-3.68	4.34
Gray	3.28	5.65	12.15	2.03

Effects of Color, Light, and Device on the Input Time

We used the completion time as a measure of perceived effort. Only the color had a significant effect on the time ($F_{3,474} = 15.7$, $p < 0.001$), but neither the device type nor the light conditions affected the time. Combined with the completion time for each color, we found that red (mean = 28.33 s, SD = 20.71) was faster than yellow (mean = 53.14 s, SD = 44.3), blue (mean = 69.66 s, SD = 62.37), and gray (mean = 52.97 s, SD = 46.26).

We can infer that yellow, blue and gray required more effort with regard to color selection than red. These results are also consistent with the post-interview results. More than 50% of the participants reported that yellow and blue were difficult to locate on the color picker, whereas no one noted any difficulty when locating red. There were no other interaction effects observed.

EXPERIMENT 2: CROWDCOLOR EVALUATION

From experiment 1, we generated two types of CrowdColor (ACC, RCC) for each color tested. We confirmed that all ACC colors were more accurate than the RCC colors except for one condition, when blue was viewed on the iPhone in the 3800K lighting environment (ΔE=-3.68, Table 2). Therefore, in the second experiment, we evaluated the best performing CrowdColor via the subjective agreement level to assess the perceived color similarity.

Method

Another group of participants was recruited (N = 18; 9 males, 9 females), whose age ranged from 19 to 31 years (mean = 23.4). They were given four color stimuli and two devices. Instead of undertaking color input tasks, they were asked to evaluate the CrowdColor generated by the previous 31 participants. We generated the best performing case by aggregating and coordinating the input data for the corresponding device and lighting conditions. The participants were asked to respond to this statement: "*I think the color of the digital swatch is identical to the object (paper) color.*" The same color stimuli were used for this evaluation, which was based on their subjective opinions on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). A total of 144 responses were collected. After completing the questionnaire, each participant was interviewed in order to assess the acceptability of CrowdColor.

Results

The results demonstrate that regardless of color or device differences, the CrowdColors were acceptable overall (>4) (Figure 3). In particular, red was rated highest among the four color stimuli. This is in agreement with the objective color distance measurement results, where red achieved a color distance of 2, which indicates that the difference was barely perceivable by the human eye.

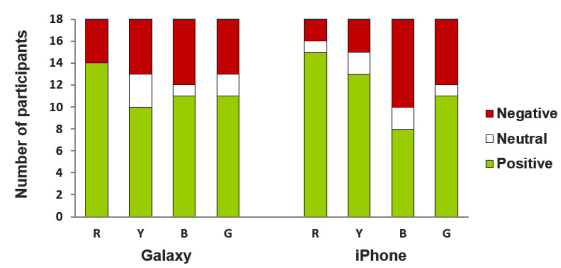

Figure 3. Subjective color agreement evaluation of red, yellow, blue and gray CrowdColors rendered on two devices

The post-experiment interviews illustrated that the participants generally agreed that the CrowdColor result was not precisely the same as the real color, but they agreed that having the CrowdColor would be very helpful when attempting to perceive the correct color when engaging in online shopping. Most participants reported that the CrowdColor result would be more trustworthy than images provided by the sellers.

DISCUSSION

Design Implications of Color Crowdsourcing

Not All Colors are Locatable

Our color-input time assessment demonstrated that the time varied for each color stimuli under different lighting conditions. The participants reported that some colors were very easy to locate, while others were not. This resulted from the limited gamut of the display, and it is unavoidable. This is a problem as well as a limitation of current display technology. Even high-end displays used by designers cannot fully display all colors. However, as display technology evolves, the accuracy of CrowdColor will naturally increase.

Increasing Reliability with Collaborative Filtering

Throughout the experiment, we noted that some of the participants expressed difficulty in matching the color to the stimuli using the color picker. Even after their selection, some doubted that their selected color was correct, whereas others were very confident in their color choices. This implies that color perception and selection abilities vary depending on the individual. Hence, the color accuracy varies depending on the user. Therefore a self-evaluative process can be used to increase the accuracy of CrowdColor. This involves giving less weight to less confident inputs and more weight to more confident inputs. Similarly, peer evaluations could weaken or strengthen the crowdsourced color inputs by way of collaborative filtering.

Applications in Real Settings and Limitations

It is arguable that CrowdColor is limited to a controlled environment only. Due to the limited crowdsourcing resources and the numerous types of devices and lighting conditions, it was challenging to conduct the study *in situ*. If the CrowdColor application is applied to large online shopping malls, the color inputs from customers will be sufficient to provide a user environment adaptive CrowdColor data in the future. As more CrowdColor inputs are accumulated, we expect that the color accuracy will increase.

We mainly considered standardized color papers as color stimuli. For generalizability, a further study of various materials, such as fabric and metal, is needed because these materials may influence color perceptions. For example, fabrics may have lower color accuracy levels and longer input times, possibly due to the more complex textures and reflections.

The average color input time was less than one minute, which is a short period of time. This supports the possibility of the use of CrowdColor in real-world applications. If color selections were added to the text review, it would be beneficial for color-sensitive customers in a color-sensitive product category. However, there is some concern over sellers who want their product colors to remain “untrue” for beautification purposes. In this case, CrowdColor does not need to work within a shopping mall; instead, a third-party color review platform that collects color reviews from many shopping malls can be used. This approach could avoid conflict with the some sellers’ intentions.

CONCLUSION

The online shopping market is growing at an exponential pace. However, inaccurate color representations online, particularly in the fashion domain, hinder customers in their purchase decisions. In this paper, we presented CrowdColor, which is a mobile crowdsourcing system that generates a representative color of a product using the crowd’s color inputs. CrowdColor was evaluated by 49 users, and it was found to be capable of providing acceptable color accuracy when used with online shopping. We assessed the effects of lighting and display types with different colors in order to demonstrate the interaction effects of these factors.

Despite the complex nature of color perception, we have taken a step forward in translating individual color perception tasks into a representative CrowdColor. Several directions for future work would be to conduct a large-scale *in-situ* crowdsourcing experiment and to apply a prototype to online shopping platforms with more advanced features, such as the automated recognition of device differences and environmental factors.

ACKNOWLEDGEMENTS

This work was supported by the ICT R&D program of MSIP/IITP under grant number B0101-15-1272.

REFERENCES

1. Farnsworth-Munsell 100 Hue Color Vision Test. http://www.color-blindness.com/fm100hue/
2. Hanrahan, P. and Lin, S. Modeling How People Extract Color Themes from Images. *CHI '13,* ACM (2013), 3101-3110.
3. Heer, J. and Bostock, M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *CHI '10*, ACM (2010), 203-212.
4. Hill, B., Roger, T., and Vorhagen, F. W. Comparative analysis of the Quantization of Color Spaces on the Basis of the CIELAB Color-difference Formula. *ACM Transactions on Graphics,* ACM (1997), 16.2: 109-154.
5. Jain, J., et al. Color Match: an Imaging based Mobile Cosmetics Advisory Service. *MobileHCI '08*, ACM (2008), 331-334.
6. Kevin R. Parker, Philip S. Nitse, and Albert S. M. Tay. The Impact of Inaccurate Color on Customer Retention and CRM. *Informing Science: the International Journal of an Emerging Transdiscipline, Volume 12,* (2009), 105-121.
7. Kim, J., et al. Illuminant Adaptive Color Reproduction based on Lightness Adaptation and Flare for Mobile Phone. *IEEE International Conference on Image Processing 2006*. IEEE (2006), 1513-1516.
8. Mantiuk, R., Daly, S., Kerofsky, L. Display adaptive tone mapping. *ACM Transactions on Graphics*. ACM (2008), 68.
9. Stone, M. A Field Guide to Digital Color. *Natick, MA. A. K. Peters, Ltd.* (2003).