Reprinted from
Journal of Imaging Science and Technology® 64(5): 050407-1–050407-16, 2020.

The Effect of Texture on Perceived Memory Color Quality

Anku Anku and Susan Farnand▲
Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, NY, USA
E-mail: axa2539@rit.edu

Abstract. *We learn the color of objects and scenes through our experience in everyday life. The colors of things that we see more frequently are defined as memory colors. These help us communicate, identify objects, detect crop ripeness or disease, evaluate the weather, and recognize emotions. Color quality has become a priority for the smartphone and camera industry. Color quality assessment (CQA) provides insight into user preference and can be put to use to improve cameras and display pipelines. The memory color of important content like human skin, food, etc. drives perceived color quality. Understanding memory color preference is critical to understanding perceived color quality. In this study, grass, sky, beach sand, green pepper, and skin were used to perform memory color assessment. Observers were asked to adjust patches with four different textures, including computed textures and real image content, according to their memory. The results show that observers adjust the image patch most consistently. In cases where the artificially generated textures closely resembled the real image content, particularly for the sky stimulus, which resembled a flat color patch, participants were able to adjust each sample more consistently to their memory color. To understand the relation between memory color and the color quality preference for camera images, a second experiment was performed. A paired comparison for familiar objects was performed with five different color quality images per object. Two of these five images were rendered from the results of the memory color assessment experiment. Additional images included were the three most preferred color quality images from a rank order CQA. This experiment was performed by naïve observers and a validation experiment was also performed by Munsell Color Science Laboratory observers. The results for color image rendering preference for each memory image content vary. The results show that for most of the colors, people prefer the top three camera color quality images used from the rank order CQA. For grass, however, the color quality preference is highest for one of the memory color assessment results. In this experiment, images rendered to reflect memory color do not match observer preference.*

[DOI: 10.2352/J.ImagingSci.Technol.2020.64.5.050407]

1. INTRODUCTION

Memory color is the color that a person recalls for familiar objects and scene content, such as grass, sky, skin, beach sand, and food items. We tend to have a very short-term memory for the color of objects, and we are not very accurate if asked to recreate them [1]. The colors of certain objects that are noticed by a person daily, such as the sky, grass, and skin, tend to be recreated with relative consistency. However, earlier results have shown that these recreations will vary for different observers not only due to differences in geographical locations, cultural backgrounds, and other such factors but also due to significant variation among observers from the same region [2–5].

Image quality is a crucial element for qualitative measurements of many consumer electronic products like televisions, cameras, displays, and printers. In the case of the smartphone, this bar is raised with the release of every new generation. Currently, cameras play a key role when consumers buy a smartphone. Customers' image quality expectations increase every year for smartphone cameras. With image quality rising steadily, smartphone processors and memory management have to keep pace. The results are improving with images being delivered at a quicker speed. Such developments have led to more flexibility with respect to image processing and integration of complex image processing techniques into existing applications. There are many factors affecting image quality such as sharpness, image noise or non-uniformity, and geometric distortion, but it is fair to say that color plays a vital role in the perceived quality of an image [6]. Colors not only spark emotions and engage a user but also decide the likability of a certain image. Popular examples are filters provided within the consumer smartphone and advanced amateur cameras and social media applications, which increase the vividness of an image. It is important to understand better the perceived image quality for pictures and to develop a procedure for evaluating them as a part of the camera development and design process.

Memory color for pictures containing sky, grass, and human skin drives the perceived color quality. The memory color of, for instance, the sky is affected by the geographical experience of an individual's memory of the sky. A person from a polluted city might not experience a clear blue sky but rather would see a very hazy sky. When presented with a blue-sky image, the person does not feel a natural connection to this perception. Other examples include the different shades of grass from dry yellowish to bright saturated green. Therefore, the geographical location and also the cultural background of an observer play an important role in understanding the memory color and hence the color preference for camera images. It is important to investigate the underlying relationship between the memory color and color quality assessment (CQA) [2]. Previous research on memory color assessment of familiar scene content, in which uniform patches were used, showed lack of agreement with the real scene content memory color [7]. In this research, the impact of different texture types on memory color is

▲ IS&T Member.

Received Mar. 30, 2020; accepted for publication July 4, 2020; published online Sep. 9, 2020. Associate Editor: Norimichi Tsumura.

1062-3701/2020/64(5)/050407/16/$25.00

explored. Furthermore, a perceived color quality experiment incorporating the results of a previous rank order color quality assessment experiment and a memory color texture assessment was performed.

This study involves two experiments. In the main experiment, we study the impact of different texture types on memory color. We used the Method of Adjustment technique in which observers were asked to adjust a textured patch according to their memory of important scene content colors. Adjustments were made in the CIELAB space because it is relatively perceptually uniform. This means that the change in visually perceived color is linear to the values in the CIELAB space [8].

Using images rendered based on the results of the previously reported rank order CQA and the memory color experiment reported here, we also conducted a color quality assessment experiment to understand the relationship between memory color and color quality preference [7]. We used a paired comparison technique for this perceptual testing in which the observers were shown two images side by side and asked to select the one that they prefer.

2. BACKGROUND

Memory color has been a topic of research for many years. One of the earliest substantial studies on memory color was done by Bartleson in 1960 [1]. In this study, he performed an experiment where observers were asked to pick their memory color of 10 different familiar objects including brick, grass, sky, and skin from Munsell Color chips. His results showed that for most of the objects, the memory color tended to be more saturated and increased in lightness. Similar results were found in the experiments performed by Adams in 1923 and Newhall and Pugh in 1947 and many more researchers in later years, including the authors' previous research [7, 9–15]. In 2015, a study showed that people tend to have saturated memory color for not only familiar objects but also for their unique personal items [16].

Further experiments were performed in 1980 to study the comparison of memory and preference for color by Siple et al. [17]. Observers adjusted the color of two sets of stimuli. One used the silhouettes of different fruits and the other used their textures on a generic shape. The results showed that the memory color and preference color were more saturated than the actual color of the fruit. The memory and preference color were independent of the shape and texture of the object. The texture for the stimuli made observers do the adjustment faster and more comfortably but did not produce any significant effect on the memory color. This study indicated that texture has influence on the overall appearance of the object itself but does not contribute to the color appearance.

Another study by Bodrogi and Tarczali [18] shows the impact of the image content on the memory color. The study included stimuli of sky, skin, and plants. The adjustments were performed on both blurry images and clear images. In contrast with the Siple study, these results showed that the image content affects the memory color of the observers. The authors noted that the skin adjustments had the least variation among the stimuli [18].

The memory color shift is more noticeable in photos than in color patches. In 2011, Smet et al. investigated the color appearance of nine familiar real objects in an LED illumination box under different illumination spectra with respect to memory colors of the objects [3, 19]. Results suggested that the observers' memory colors of the objects were of higher chroma when compared with the actual real object color measurement. A study by Olkkonen et al. [20] shows that the effect of memory color on the color appearance of fruits and vegetables is affected by the objects' shape, texture, and illumination [20]. The study used photographs of 3D fruit shapes with no texture and 2D outline shape and showed that the effect is stronger when stimuli are more realistically rendered as compared to when they are not. A similar study was performed by Vurro et al. [21], where they studied the effects of surface texture, 3D, and contour diagnosticity for fruits. The results show that as we increase the naturalness of the stimuli with the surface texture or the shape of the fruit, the accuracy and precision of the memory color increase [21].

Vurro et al. [21] study uses texture stimuli that resemble the object's original texture without highlights and shadows. Our study is the first study, to our knowledge, that includes different degrees of artificially created textures as an independent variable to assess the memory color for familiar objects. Our study focuses on understanding how we relate the degree of fidelity of texture to the memory color of familiar objects like grass, beach sand, sky, skin tone, and green pepper.

Memory color effect has been widely studied in the field of color constancy, where researchers have used familiar object colors to obtain color constancy for unfamiliar objects in the scene. Hering stated that the memory colors of familiar objects could have a potential role in causing colors to remain constant under changes in illumination [22]. The perceived color of an object is affected by the prior knowledge of an object's color. The results have also shown that color constancy was higher for scenes containing familiar objects. The yellow color of a banana interferes with the color constancy and also with the perception of real color [12, 13, 23].

A recent study was performed by Yan and Suk on one of the familiar image scene contents discussed in the other studies, skin [24–27]. The study investigated the quality of three different image renderings based on memory color, measured skin tone, and digitally preferred skin tone. Observers were asked to rate five parameters: naturalness, realism, emotion, appropriateness, and preference. The results showed that stimulus images rendered based on the memory color corresponded to the emotional parameter. Stimuli rendered based on measured skin tone appeared more natural and real to the observers. Finally, the images rendered based on the preferred skin tone adjustment corresponded to the most preferred images. This result shows that the memory color and preferred color for skin tone

Figure 1. Original images showing various common objects in typical scenes with different types of lighting. These images are cropped to create the stimuli for the experiments.

are not the same. Another aim of our study is to evaluate the preference of two different image renderings based on memory color and preferred color quality for familiar scene contents. This study will help clarify the relationship between memory and preference color.

In previous research conducted by the authors, a color quality assessment was performed on images taken using 20 devices retailed in 2014 having cameras that spanned a range of specified capabilities [7]. The original scenes used for the study were taken in various day-to-day scenarios typical of smartphone camera use [6]. Results were reported on the individual image quality characteristics relative to overall quality as well as to the objective measures of each characteristic. A high correlation was found between the individual characteristics—color quality, tone quality, noise, and sharpness—and the overall quality of the images. In comparing objective and subjective results, there was strong agreement between the visual results and objective measurements for noise and sharpness. However, the objective measures did not provide predictive estimates of perceived color quality, which was likely due to the more subjective nature of color [6]. In follow-on testing conducted to determine whether patches containing memory colors can be used to predict the color quality of the entire image, these original images were cropped to create 200 × 200 pixel patches of familiar scene content including sky, grass, beach sand, foliage, wood, skin, and brick (Figure 1). The color quality of these images was assessed in rank order perceptual testing. The observers were asked to arrange the images from the most preferred to the least preferred based on their color quality preference. The results showed that camera preference varied with image content. This indicates that the context of the scene is necessary when evaluating the overall camera color quality [7] .

The validation of the above experiment was carried out using anchored scaling perceptual testing [7]. This testing technique uses lower and higher anchor images to which the test images are compared. The lowest-ranked image in the rank order color quality perceptual testing was used as the low anchor and the highest ranked as the high anchor. The observers were asked to give a score to the test image based on the low anchor and high anchor scores. For further information on perceptual scaling experimental protocols, see [28]. The results showed that the anchored scaling perceptual tests were generally consistent with the results of the CQA. The rank order CQA results were also compared with the Farnand et al. (2016) [6] anchor scaling CQA results, which used the full-context images as test stimuli whereas the rank order CQA by Anku et al. [7] used only patches from these images [7]. The comparison results show a high correlation for most of the image patches.

Figure 2. Patches selected from the original images.

In addition to the rank order CQA, a memory color assessment was performed by the same observers. A uniform patch was shown, and the observers were asked to adjust the patch color according to their memory of the color of familiar things, including grass, sky, face, beach sand, and vegetables (Figure 2), using three sliders: lightness scale, red–green scale, and yellow–blue scale in CIELAB space. The sixth highest ranked cropped patch from the rank order color quality assessment was used as the starting point. The results show that observers tend to choose saturated colors, in agreement with previous research, and may not be accounting for noise affecting the object appearance. For many observers, the adjustment of uniform patches to represent memory colors is consistent in hue and saturation. Furthermore, the average pixel color values in CIELAB space of the top three ranked images of rank order CQA were compared with the adjusted CIELAB values for the memory color assessment. The results show that observers remember the color as more saturated. It is possible that observers are not accounting for factors involved in object appearance like dirt and shadows in the grass, dust particles in beach sand, melanin spots in skin, and clouds in the sky. The results indicate that rendering preference is not well predicted just by memory adjustment of uniform patches. The color quality of familiar scene content was fair when rendered by the memory of uniform patches but could have been improved if the rendering mechanism was developed by methods taking into account the texture characteristics.

This leads to the question of how adding texture to uniform patches would impact the perception of memory color. To evaluate the effect of non-uniformity, we performed an experiment using textured patches of important memory colors along with uniform patches of these colors.

3. METHODOLOGY

This study is performed in two parts. The first part is aimed at memory color assessment for familiar image scene content with four different texture levels, while the second part is aimed at color quality assessment. The same original image scenes from the previous experiments are used. The pictures include familiar scenes like beaches, parks, restaurants, and food. Lighting conditions and framing are kept constant across all pictures [6]. The set of images is shown in Fig. 1. The contents of these images are elements common in everyday photography such as the sky, grass, sand, vegetables, and skin.

Figure 3. Experiment 1 setup, which consists of a display showing a GUI consisting of a textured familiar object patch. The observer adjusts the L^*, a^*, and b^* sliders according to their memory color.

To evaluate the effect of texture on memory color assessment, we created color patches with four different texture levels (Figure 5). The textures in these patches were chosen to increase at distinct, although discrete, levels, starting from a uniform color patch on the left to Texture 1, which is fine-grained, derived from the sandstone texture property of Adobe Photoshop®. Texture 2 is a coarse-grained texture created using Adobe Photoshop® and was made to resemble burlap, and we included a cropped section of the original image representing the real-world scene content.

For the main experiment, these four different texture levels for five familiar objects were used as stimuli. For the second experiment, we showed observers five different color renderings for each of the five familiar objects.

4. EXPERIMENTAL SETUP

Experiment 1, which examined the memory color for common objects and scene content in four different texture types, was performed in the Perception Laboratory at the Munsell Color Science Laboratory with the setup shown in Figure 3. A MacBook Pro laptop was connected to the Eizo CG248 color display, which was used to present the stimuli to the observers. The display was calibrated using a Photo

Figure 4. Experiment 2 setup during Imagine RIT, which consists of a black sheet to avoid any interference of light. The GUI consists of two familiar object patches side by side, where the observer picks the one they prefer based on color quality.

Research Spectrometer PR-655 [29]. The wall behind the monitor was painted gray (Munsell N5) and was illuminated by a D65 metal halide lamp filtered to have a luminance level equivalent to that of the average image rendered on the display, 70 cd/m^2. This setting is recommended by CPIQ standards to reduce observer fatigue compared to a completely darkened surround [30].

The display and the experiment lab light source were switched on 30 minutes before the experiment was started to maintain uniform illumination. Other lights in the laboratory were turned off to avoid stray light. The table on which the display was placed was covered with a gray sheet of paper to make the observer's field of view as uniform as possible. The observers were seated 85 cm from the display. A MATLAB driven Graphical User Interface (GUI) was used to present the stimuli and the adjustment sliders on the display, enabling users to adjust the patches to their preferences. The background of the GUI was set to a similar gray to that of the wall and the table sheet.

Experiment 2 was carried out using a similar setup (Figure 4) to that of Experiment 1 except that it was conducted during Imagine RIT 2019, which is a technical festival that takes place at the Rochester Institute of Technology, which is typically attended by about 30,000 people from a wide range of cultural backgrounds, ages, and professions. We wanted to take this opportunity to show people what we do by encouraging them to take part in the study. The experiment was conducted in the corner of a large room where a black cardboard was placed behind the display, as shown in Fig. 4, to avoid glare and reflections. The room lights were turned off, but there was some illumination from daylight. The luminance level was 67 cd/m^2, which is similar to the Experiment 1 setup.

5. OBSERVERS

All the observers were tested for color vision using the Ishihara plate test. There were a total of 30 observers for Experiment 1, among whom 10 were female and 20 were male. All the observers had normal color vision and normal or corrected to normal visual acuity. Twenty-five observers had some color or imaging science background, whereas the other five did not. The age range for most of the observers was between 20 and 30.

For Experiment 2, 37 people participated including 17 females and 20 males. All the observers were naïve observers and did not have any color or imaging science background. The observers were 15 to 50 years of age. These observers were visitors to the *Imagine RIT* festival. A consent form was signed by all the observers and for observers younger than 18 years of age, their parents' consent was obtained. To validate Experiment 2, the same experiment was performed by the Munsell Color Science Laboratory students and faculty. Twelve observers participated in the validation process. The nationalities of the observers were American, Indian, Asian, Iranian, and European. The observers were diverse in age, gender, and nationality in order to better encompass a wider audience of people. All the experiments were approved by the RIT Institutional Review Board (IRB) for Human Subjects Research.

6. PROCEDURE

6.1 *Memory Color Assessment*

The main aim of this experiment is to investigate the effect of texture on memory color preferences for familiar objects. The elements found in many of the pictures we take on a daily basis have textures and details in them that may affect perceived memory color. In the images used in our study, these details could be dirt in the grass, clouds in the sky, freckles and hair follicles on skin, damp or wet sand, etc. We want to study how observers take these details into account when recollecting their memory color of these common elements by exploring how introducing textures of various levels may change the memory color recollection from that of a uniform color patch.

The stimulus images (Fig. 2) featured familiar scene content including human face, grass, green pepper, beach sand, and sky. There were five familiar elements with four different texture levels, resulting in a total of 20 images as shown in Fig. 5. Observers were presented with these 20 images, one at a time, in randomized order to avoid bias.

The observers' task was to adjust the image according to their best memory of the color of the respective content. The layout of the experimental interface is shown in Figure 6. There are three sliders: L^*, which is dark to light (left to right); a^*, which goes from greenish to reddish (left to right); and b^*, which goes from bluish to yellowish (left to right). The intervals of the L^* slider bar were linearly spaced with a step size of 0.1, and the a^* and b^* slider bars were similarly linear with a step size of 0.039. Observers were advised to use the slider button on the extreme ends of each slider bar for precise and controlled adjustments. The UI sliders change

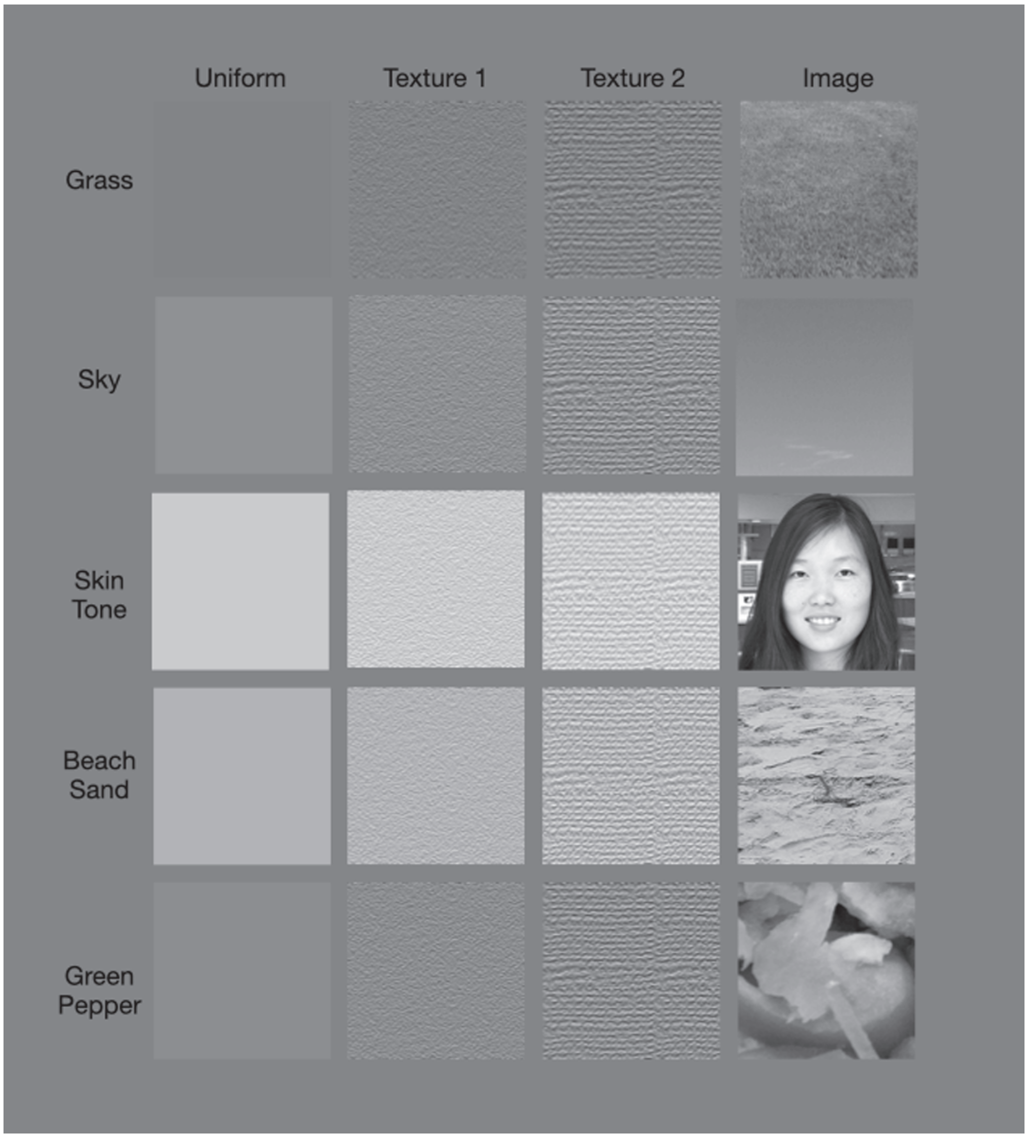

Figure 5. Image set for Experiment 1 with different levels of texture. The base color for uniform patch, Texture 1, and Texture 2 was calculated by the average of the image patch.

the respective attribute on each pixel on the image patch, so any adjustments will be reflected on the entire image. A short text description for each stimulus was presented on the top of the GUI. No additional information was provided about the scenes. This procedure is common in previous studies involving memory color [1, 7]. The observers were adapted to the room lighting while reading the instructions, which took about 5 minutes. The average time taken for this experiment was 30 minutes.

6.2 *Color Quality Assessment*

In this experiment, we evaluated the preference for color quality for familiar scene content including grass, sky, beach sand, skin tone, and green pepper. There were five objects, each with five different color renderings. Therefore, the total image set comprised 25 images as shown in Figure 7. Among these five different color renderings, three were selected from the authors' previous research on rank order color quality assessment [7]. The other two renderings, two medoids, were selected from the 30 observers' data set from the memory color preference experiment. The process for calculating these medoids was as follows. The outliers were excluded using the MATLAB outlier function value; 2-mean clusters for the textured adjusted points were calculated along with the medoids for these two clusters. These two points were then used to color-render two stimulus images. In Fig. 7, the rank 1, rank 2, and rank 3 columns indicate that these images are ranked #1 (highest preference), #2, and #3, respectively, in the authors' previous rank order CQA. Medoid 1 and medoid 2 represent the images that were color-rendered based on memory color assessment.

This experiment used a paired comparison protocol with a total of 50 pairs. The GUI displayed two images side by side as shown in Figure 8. The order of these pairs was

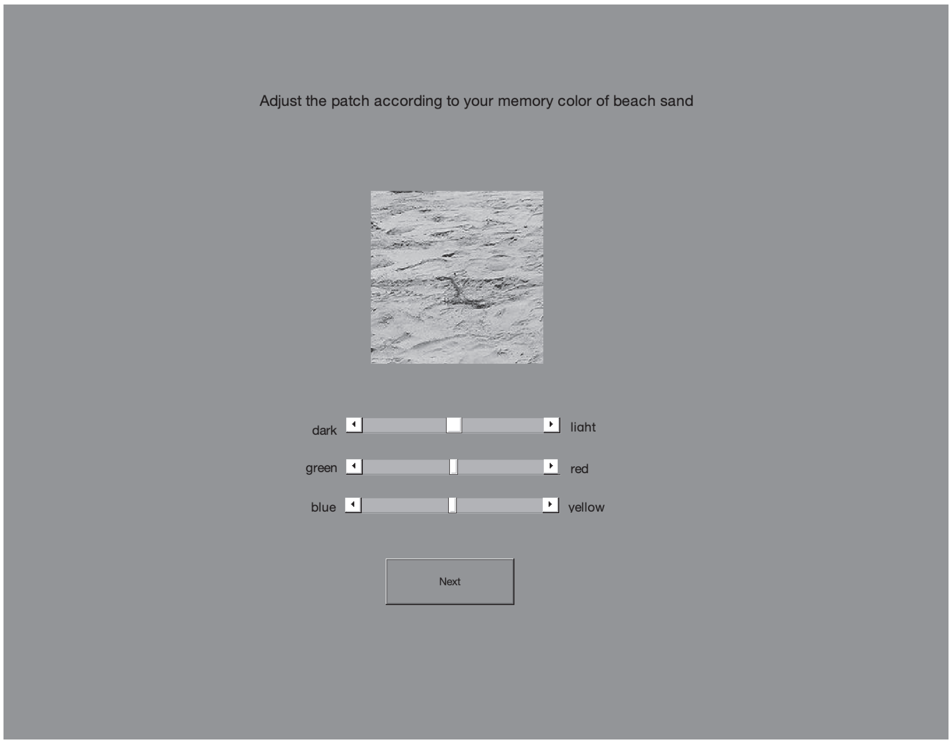

Figure 6. GUI for memory color assessment. It displays texture patches of common objects. The observers are asked to adjust them according to their memory color of these objects using sliders. Slider 1 is a lightness scale, slider 2 is a^* of CIELAB, and slider 3 is b^* of CIELAB.

randomized to avoid bias. The observers selected their color quality preference using the left and right arrow keys, which corresponded to the left and right images. Observers adapted to the room lighting while receiving instructions, which took about 5 minutes. The average time taken for this experiment was 10 minutes. A validation test was performed using the same procedure as that for the the main experiment except for the slight variations in the experimental setup discussed above.

7. RESULTS

7.1 *Experiment 1—Memory Color Assessment*

For the memory color assessment, we showed the observers four different textured patches of familiar image scene contents and asked the observers to adjust them according to their memory of the color of the scene content. From the previous research results, it has been observed that uniform patches alone are not enough for observers to recreate the color [7]. It was hypothesized that additional information about the object that texture provides would help observers recreate colors more precisely, leading to increased observer consistency. It was also expected that the recreation of memory color would vary as the observers are influenced by factors such as geographical and cultural backgrounds [2].

Figure 9 shows the four different texture types for grass used in Figures 10 and 11, which present the results for the memory color of grass. The plot in Fig. 10 has a^* values on the x-axis and b^* values on the y-axis. The ellipses are the 90% confidence interval for the respective textures. The data points represent the observers' adjusted L^*, a^*, and b^* values.

Fig. 10 results show that the adjusted memory color for uniform patch spreads from −90 to −20 in a^* and 20 to 60 in b^*. For Texture 1, which is a sandstone texture, the ranges are narrower, with an a^* range of −50 to −20 and a b^* range of 20–50. A similarity in appearance is observed between Texture 1 and grass. Texture 2 has a burlap structure, which is dense and differs in appearance from grass. The range is spread out more for a^* and b^* from −80 to −20 and 25 to 55, respectively. The image patch, which is an image of actual grass, has the most consistent result. The a^* ranges from −40 to −20 and b^* ranges from 25 to 50. From appearance, we can see that Texture 1 is more similar to the image patch and that may, therefore, be the reason why its ellipse is more like that for the image patch. Note that if the two outliers on the left were removed, the ellipse for Texture 2 would also be fairly similar to that of the image patch, meaning both textures produce results more like the image than does the uniform patch. In addition, note that the image patch ellipse is a subset of all the other four textures, which means that all the points within the magenta ellipse can also represent the other three.

Fig. 11 shows a pattern similar to that described for Fig. 10. Overall, the lightness scale ranges from 30 to 60, whereas the chroma scale has a wider range from 30 to 90. Overall, the observers' data points are more clustered toward 40–60 for chroma and 40–50 for lightness. The uniform patch and Texture 2 are the most scattered, whereas Texture 1 and the image patch are less scattered and more consistent,

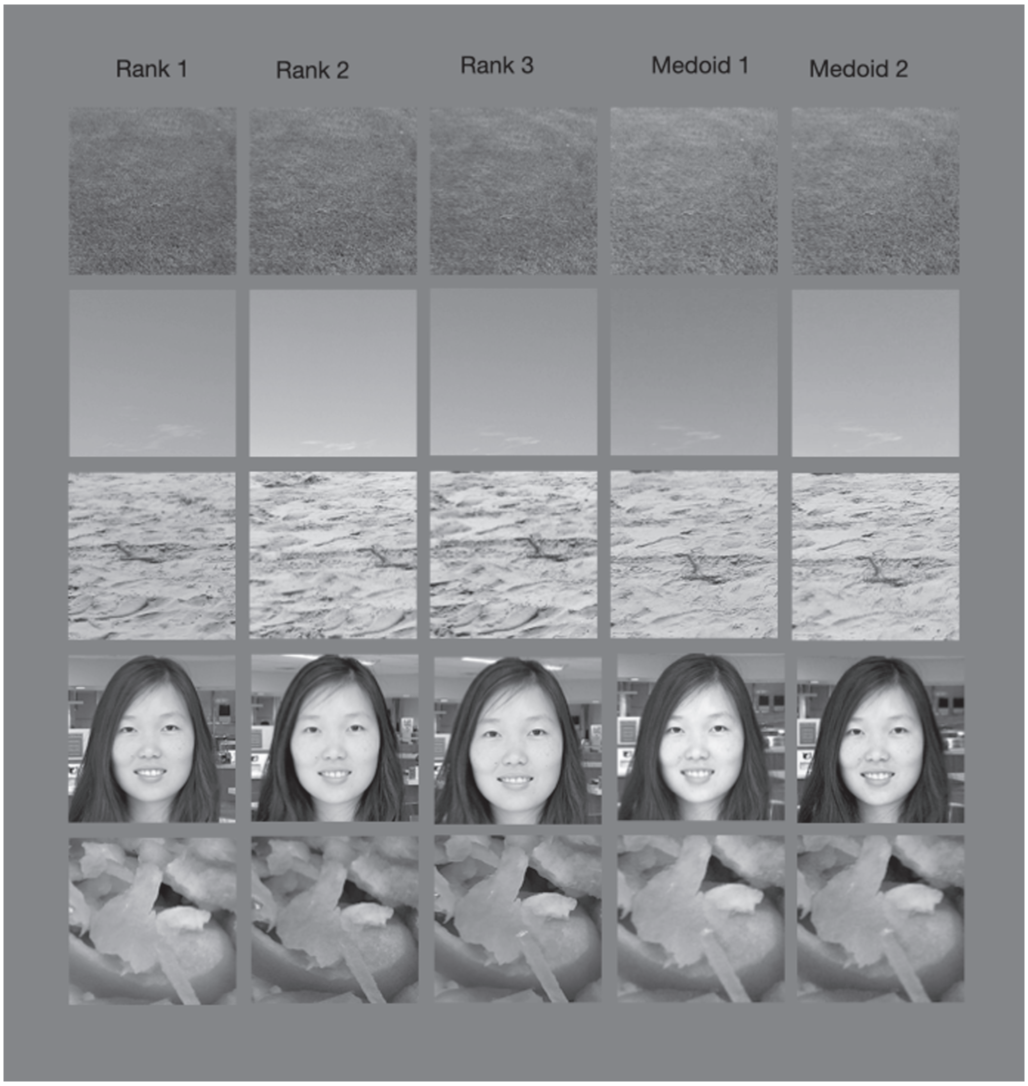

Figure 7. Image set for Experiment 2, color quality assessment. Rank 1, rank 2, and rank 3 are the top three selected images from the rank order experiment. Medoid 1 and medoid 2 are the two classes that represent the result of the memory color assessment experiment.

indicating that Texture 1 has an effect similar to that of the image patch.

Figure 12 shows the four different texture types for sky used in Figures 13 and 14, which present the results for the memory color of sky. We can see that the overall results for sky are more consistent than those for grass. For the uniform patch, a^* ranges from −10 to 10 and b^* ranges from −70 to −30. For Texture 1, a^* ranges from −23 to 10 and b^* ranges from −70 to −30. Texture 2 is the most scattered with a^* ranging from −30 to 20 and b^* ranging from −70 to −35. The image patch is the most consistent patch.

We can observe that the uniform patch and the image patch are similar in appearance except for the clouds in the image patch, which is the reason why the red and magenta ellipses are closer to each other. It is hard to imagine the sky with any form of texture because it is not natural to see a textured sky, which likely explains why it was harder for observers to adjust the memory color patches for Textures 1 and 2. The image patch ellipse is a subset of the other four textures.

Fig. 14 shows a consistent chroma level for all texture types ranging from about 22–62. The lightness level ranges from 50 to 90. We see that the image patch ellipse is again the most consistent one. The intersection of all four textures consists of the majority of the data points. The lightness for the uniform patch is more scattered because a uniform patch puts no limit on the memory color of the observers, which means observers can imagine the sky on a wide scale of lightness levels. We can notice that as the lightness values increase, the chroma values decrease. We have seen the same results in previous research [7]. A similar chroma line pattern is followed by all textures except the image patch, which is an actual image of the sky with a cloud. The image patch does limit the memory color because of the presence of the cloud. Increasing the overall lightness of the image patch would affect the appearance of the cloud in the image by washing it out and making it look unnatural, which

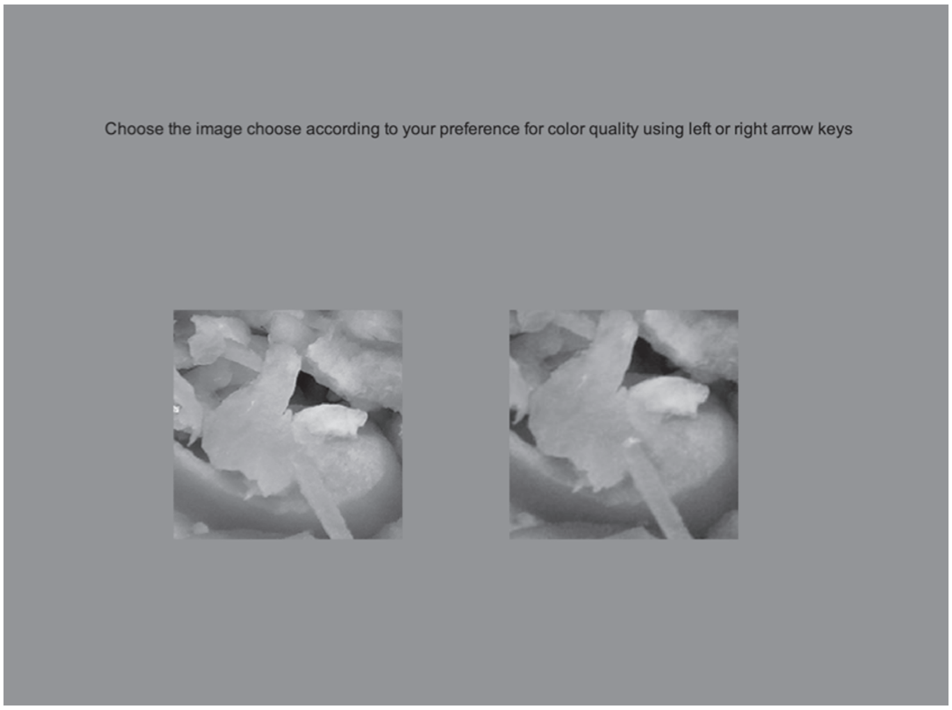

Figure 8. GUI for the color quality assessment conducted using a paired comparison between two side-by-side images. The observers were asked to pick the image they preferred based on color quality.

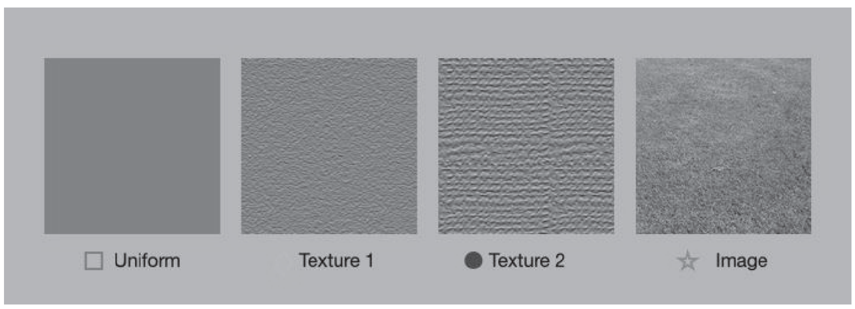

Figure 9. Texture types of grass.

essentially discourages the user from adjusting the lightness further.

Figure 15 shows the four different texture types for skin tone used in Figures 16 and 17, which present the results for the memory color of skin tone which has always been an important memory color due to geographical and cultural differences. Skin tone is the most dominant element among other familiar objects in a scene [31, 32]. The skin image patch is much more different from the uniform and texture patches than the other memory colors evaluated. It is important to note that the image patch data points are the average of only the face area, excluding the lips, hair, eyes, etc. It is challenging to imagine just a skin textured patch without the context of the body parts. This is the reason why the magenta ellipse overlaps somewhat less with the other texture types. We can see that the overall a^* range is quite consistent, which is from 2.5 to 22.5 and b^* is from 0 to 40. This is the most consistent memory color among all that were tested, which is in agreement with previous studies [7]. It is understandable for observers to recreate the color of skin with precise hue due to its familiarity. The majority of the data points for Textures 1 and 2 are within the image patch ellipse. Texture 1 seems the closest match to the actual skin areas of the image patch. Texture 2 has more points scattered, which is likely because the texture is much harder to imagine than is skin.

More variation can be seen in the lightness level relative to the a^* and b^* variations (Fig. 17). The overall range for lightness is 60–95 units. The chroma for most of the data points is in the range of 12–32. A majority of the data points are within the intersection of all texture ellipses, indicating that there is good agreement irrespective of the texture type. The Texture 1 ellipse is closest to that for the image patch, indicating that the sandstone texture is the best representative of skin among the three non-image textures tested.

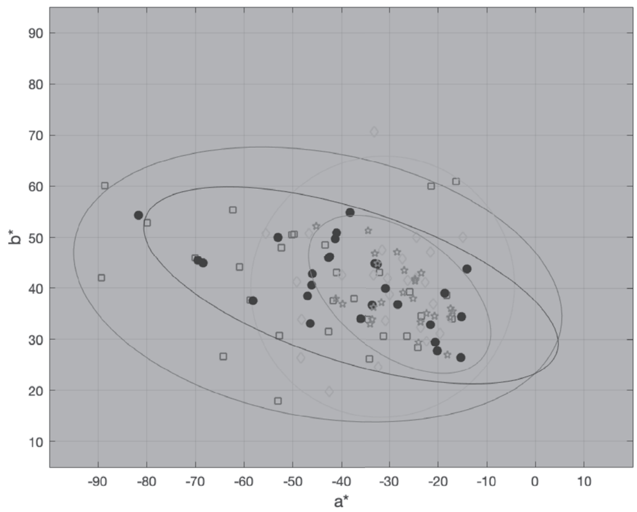

Figure 10. Memory color of grass. The x-axis is the a^* channel and the y-axis is the b^* channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

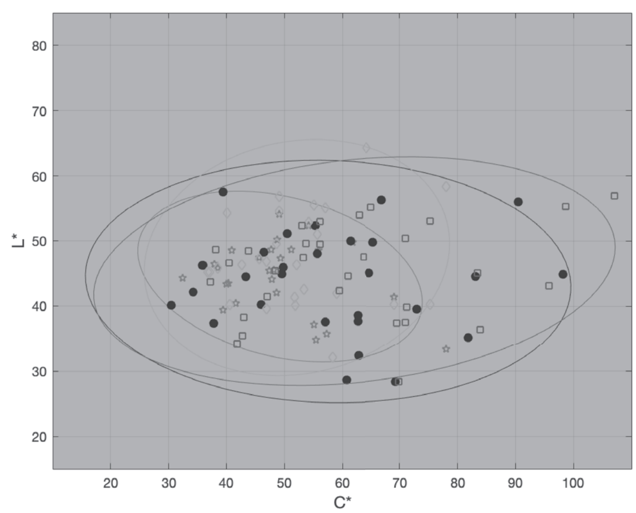

Figure 11. Memory color of grass. The x-axis is the chroma, C^* channel, and the y-axis is the lightness, L^* channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

Figure 18 shows the four different texture types for green pepper used in Figures 19, 20 and 21, which present the results for the memory color of green pepper. This suggests that observers have a less consistent idea of the color of green pepper, possibly because this color varies around the world. The a^* data points range from saturated artificial green pepper to pickled green pepper. The image patch of sliced green pepper has a large range for a^* from −60 to −15 and for b^* from 30 to 60. The magenta ellipse is not a complete subset of the other three texture types, possibly because it includes the white internals of pepper, which are ignored while adjusting the memory color of pepper for the other three textures. The intersection of all four patch types contains most of the data points, ranging from −60 to −20 in a^* and from 30 to 50 in b^*.

Lightness and chroma have a wide range for green pepper as seen in Fig. 20. Texture 1 has a smaller ellipse than Texture 2 and the uniform patch, which means that it is easier to imagine pepper while adjusting memory color. As shown in Fig. 20, the image patch data points have higher lightness values. To evaluate the effect of white pepper parts on the memory color result, the average color was calculated for only the outer skin of green pepper (Fig. 21). Interestingly, the image patch ellipse overlaps with the other three textures. Moreover, the lightness value decreases and the chroma increases. These result in a higher similarity to observers' memory of the color of green pepper, indicating that the white parts are not averaged in human memory while adjusting the pepper texture patches.

Figure 22 shows the four different texture types for beach sand used in Figures 23 and 24, which present the results for the memory color of beach sand. All the ellipses have similar shapes and sizes, which is likely because it is easy to imagine textured beach sand. All the data points overlap with each other. The uniform patch has a slightly greater range on a^* and b^* as compared to the other textures because it is a uniform patch and sand is not uniform. The b^* ranges quite a lot for all textures, possibly because of the geographical differences among the observers. The b^* range varies from 0 to 60 due to different time, weather, and season experiences. During winters, sand appears to be colder (bluish), whereas during summers, it appears to be warmer (yellowish).

In Fig. 24, the ellipses generally overlap. The lightness scale ranges from 60 to 90, and chroma has a wide range from 0 to 70. It is interesting to note that the results are more consistent for Texture 1 than the image patch. This is the only familiar object for which one of the non-image textures has a smaller ellipse than the image patch. The experimental results indicate that the image patch generally has the lowest variability in the observers' selected memory colors and that the texture patch that most closely resembles the image content has the next lowest variability.

8. SUMMARY OF EXPERIMENT 1—MEMORY COLOR ADJUSTMENT

The first experiment examined the impact of texture on the memory color of familiar scene contents: grass, sky, human skin, green pepper, and beach sand. For the memory color assessment of grass, Texture 1 and image patch ellipses are of similar shape and orientation, which suggests that Texture 1 provides a more natural appearance of grass than the uniform patch and Texture 2. Using Texture 1, observers were more consistent in adjusting the image to their memory color of grass than with the other textures.

Sky is relatively uniform and when using any texture, observers performed worse in their adjustments to memory color. For the task of adjusting the color of the stimuli to their memory color of the sky, observers performed best with the uniform patch and image patch. The average lightness

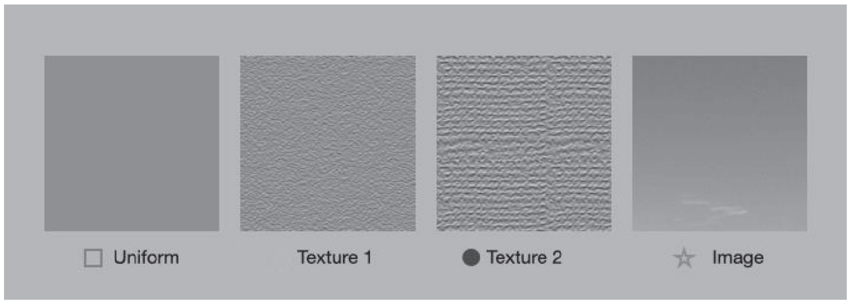

Figure 12. Texture types of sky.

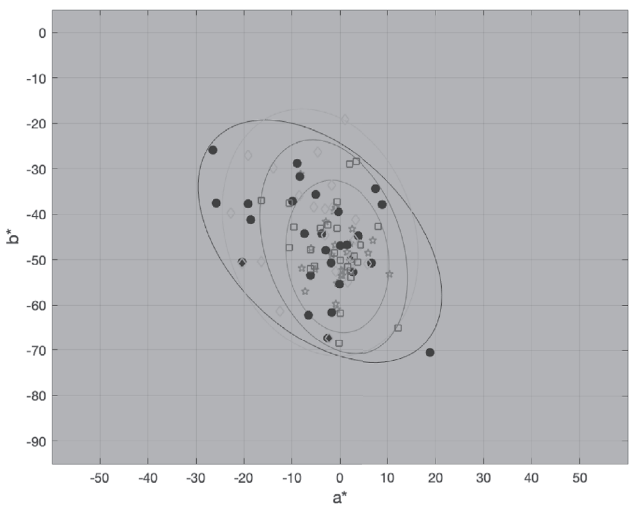

Figure 13. Memory color of sky. The x-axis is the a^* channel and the y-axis is the b^* channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

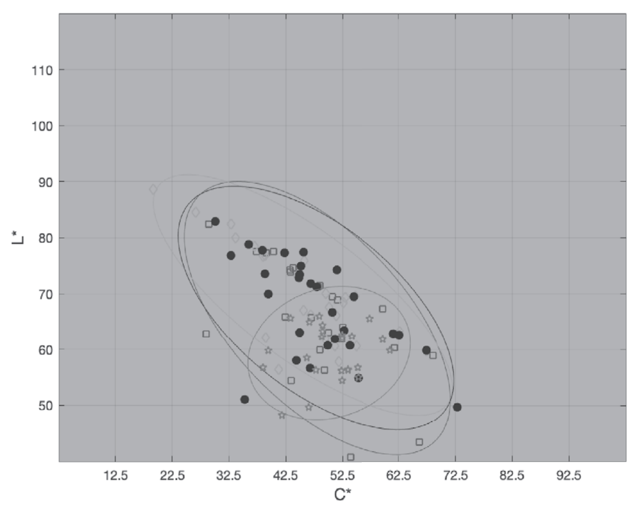

Figure 14. Memory color of sky. The x-axis is the chroma, C^* channel, and the y-axis is the lightness, L^* channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

difference in uniform patch and image patch is affected by the cloud in the image patch, which is not included in the uniform patch. When the cloud was removed from the sky image patch during data analysis, the lightness value decreased for all the data points and had increased correlation with the uniform patch data. Perhaps if a stimulus based on the uniform patch but with some cloud texture was used, then users would adjust this stimulus to their memory color for sky even more accurately.

For the skin tone adjustment task, the image patch results have the least variation. Among all of the image stimuli, adjustments for skin color using the real image patch are the most consistent. Each of the artificial stimuli was much harder for observers to adjust to their memory color for skin. Texture 1 is the second most consistent type, and the ellipse intersects more with that of the image patch, which implies that Texture 1 is closer to real skin texture.

For green pepper, the results are the most inconsistent and scattered among all the memory colors tested. However, it is interesting to note that after excluding the inner skin of pepper, the L^* and C^* ranges of the image patch line up with that of the other textures and the ellipses overlap. This indicates that when observers were asked to recreate the green pepper color, they do not average the inner and outer skins of pepper but rather just think about the outer skin.

The beach sand has a similar shape and size for all the four texture types, which is likely because it is easy to imagine textured beach sand. The uniform patch data points are slightly more spread out compared to other textures because sand is not uniform. The a^* range is very consistent for the beach sand, whereas the b^* range varies depending on the observers' memory of sunny or cloudy beach sand. It would be interesting to assess the memory color by giving more information to observers like the season, weather, time of day, etc.

8.1 *Experiment 2—Color Quality Assessment*

The results of the second experiment, in which perceived color quality using stimuli rendered from the results of the earlier rank order CQA and the memory color assessment of Experiment 1 were evaluated, are shown in Figure 25. The x-axis is labeled with the familiar objects, and the y-axis is the scale value, which is calculated using Thurstonian analysis [33]. The error bar represents a standard deviation of 0.086. The preference order is converted to a frequency matrix, which is then divided by the total number of

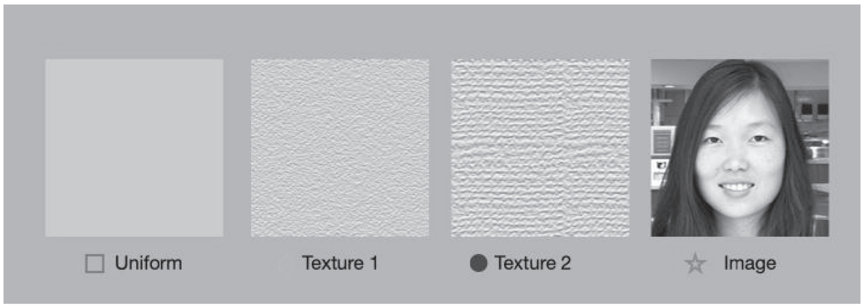

Figure 15. Texture type for skin tone.

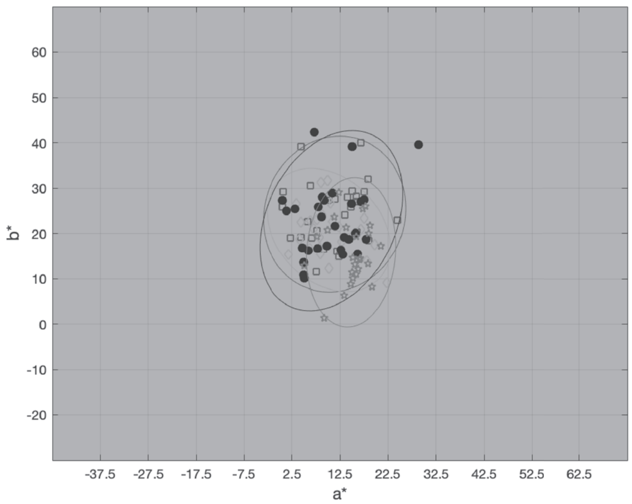

Figure 16. Memory color of skin tone. The x-axis is the *a** channel and the y-axis is the *b** channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

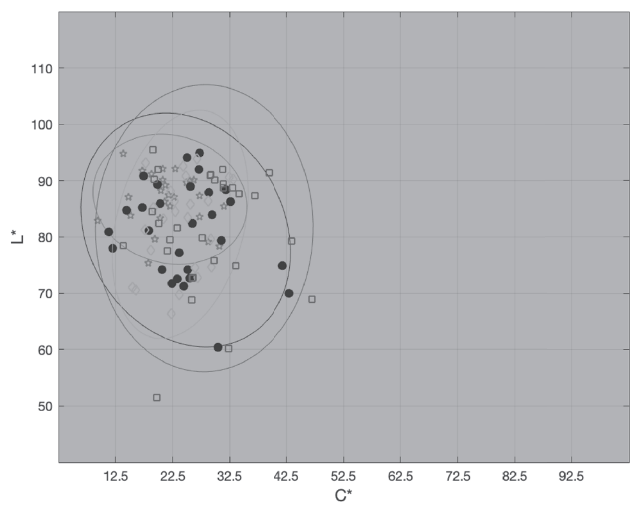

Figure 17. Memory color of skin tone. The x-axis is the chroma, *C** channel, and the y-axis is the lightness, *L** channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

observers to obtain the proportion matrix. Following this, the z-score values are calculated for the proportion matrix. These z-score values, here referred to as scale values, are normalized to avoid dealing with negative numbers.

The bars represent the five different images used as stimuli. As the legend describes, blue is the top-ranked image from rank order CQA, orange is the second most preferred, and yellow is the third preferred. Purple is medoid 1 of the first cluster of the adjusted memory color results and green is medoid 2 of the second cluster.

As we see in Fig. 25, for grass, medoid 1 is the most preferred. Fig. 7 shows that medoid 1 is lighter and has a higher chroma than the others. This means that an observer's memory color and color quality preference are the same for grass. Humans tend to discount shadows, textures, etc. and remember colors to be more saturated, and they prefer the same [1]. The scale values among the rest of the images are similar, which means that observers were not sure about their choices when comparing the other four images.

For sky, rank 1 and medoid 1, which are more saturated than the other three renditions, have higher scale values. Ranks 2 and 3 have similar scale values, which suggests that the observers like them equally. Medoid 2 is the least preferred one as it has a slight purple tint to it. Previous research has shown that people dislike purplish sky [7].

For the beach sand, the rank 2 image, the rank 3 image, and the medoid 2 image are preferred to the rank 1 and medoid 1 images. These top three preferred images all have a sunny warmth to them as shown in Fig. 7. Also, we can see in the original image of the beach sand in Fig. 1 that the sky is slightly cloudy, but the weather is mostly warm. The results indicate that the observers do not like the grayish tint to beach sand, which is seen with the rank 1 and medoid 1 lower scale values. Beach sand may be mostly imagined with sunny, warm weather. It is rare for people to visit a beach when it is raining, which implies that there is a rare memory of a rainy beach.

The color quality and memory color for skin do not agree. Our results indicate that observers like blemish-free, rosy pictures of faces, whereas their memory color tends to be simpler and warmer. The rank 3 image is the least preferred one. Observers prefer the rank 1 patch the most. The rank 2 and medoid 1 and 2 patches have similar scale values, which means that observers had a hard time deciding the best among these three.

Similar to skin, for green pepper, the color quality preferences are different from the memory color assessment.

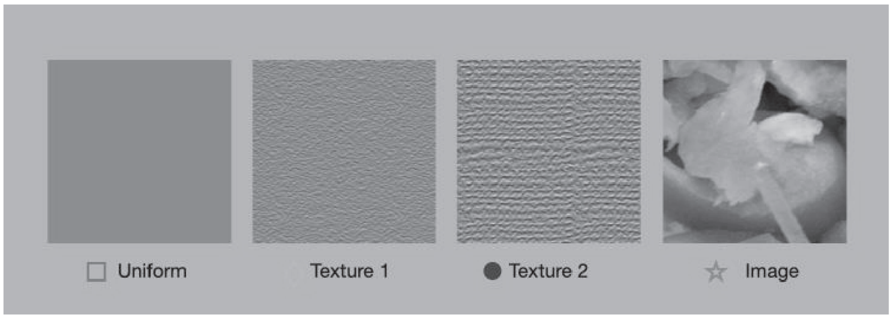

Figure 18. Texture type for green pepper.

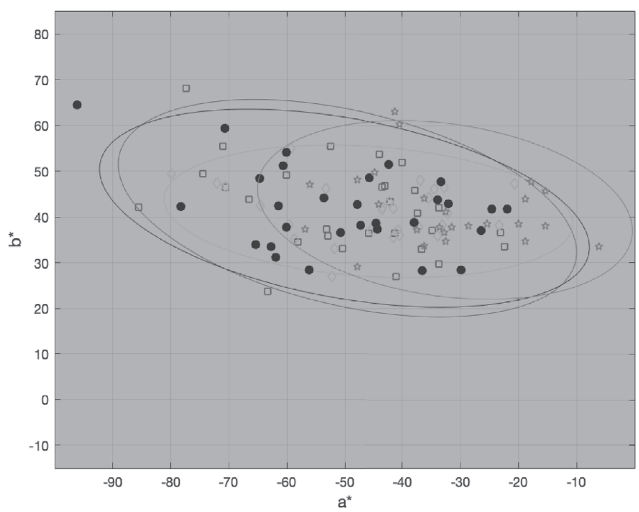

Figure 19. Memory color of green pepper. The x-axis is the *a** channel and the y-axis is the *b** channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

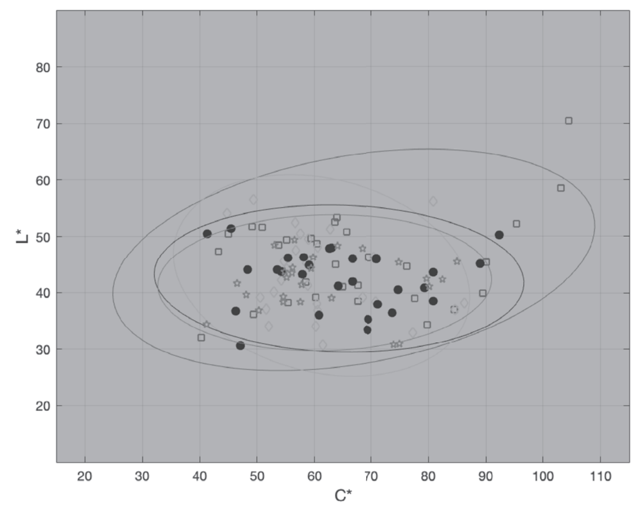

Figure 21. Memory color of the outer skin of green pepper. The x-axis is the chroma, *C** channel, and the y-axis is the lightness, *L** channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

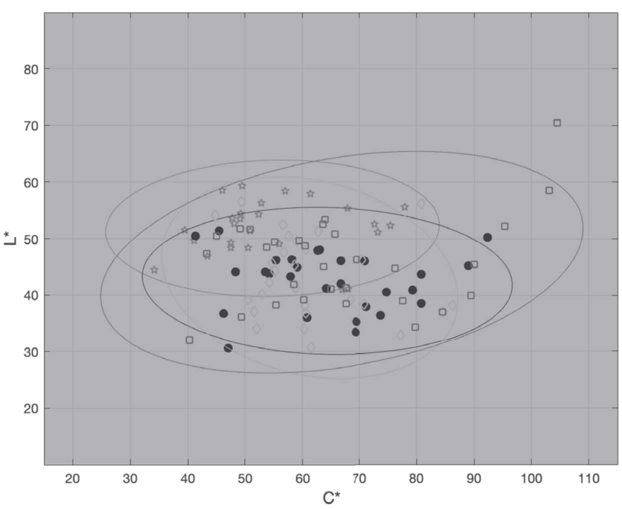

Figure 20. Memory color of green pepper. The *x*-axis is the chroma, *C** channel, and the *y*-axis is the lightness, *L** channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

Performing the color quality assessment was an easier task for the observers than the memory color assessment. Medoid 1 and 2 images show observers' memory of pepper color, whereas ranks 1, 3, and 2 are what observers would prefer the image to look like. There is a huge difference in the scale values for the CQA images and memory color images.

In Figure 26, the x-axis represents the method used to render the different color quality images, and the y-axis is the scale value. We can see that medoid 1 is effective for grass and rank 3 for green pepper. Rank 1 is highly preferred for green pepper, sky, and skin tone.

Figures 27 and 28 show the results of the validation of Experiment 2 performed by Munsell Color Science Laboratory observers. The error bars represent a standard deviation of 0.162. We can see that the overall scale values are higher, which means that there is good agreement among the observers on their preferences. For grass, the most and the least preferred are the same as those in Experiment 2, but there are some differences among the other three images. The naïve users did not distinguish between these three images, whereas the Color Science observers preferred medoid 2, a memory color image, which is more realistic, to the rank order images.

The same pattern is followed for the sky image as well. Beach sand preference aligns for the naïve and Color Science

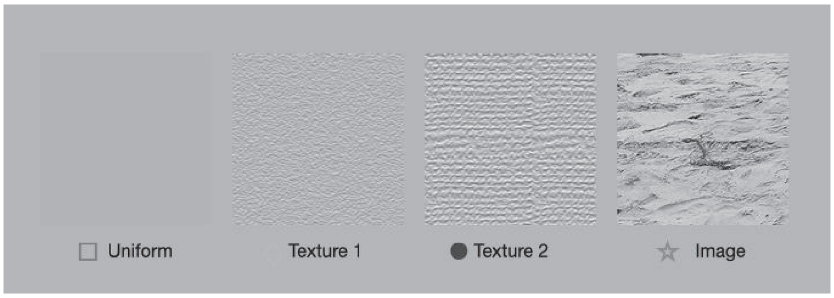

Figure 22. Texture type for beach sand.

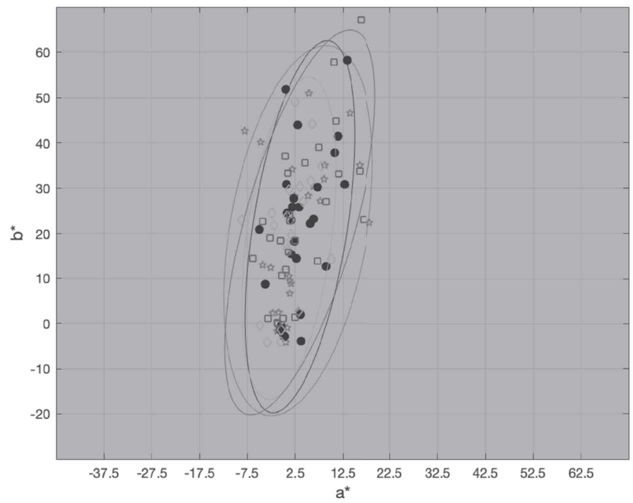

Figure 23. Memory color of beach sand. The x-axis is the a^* channel and the y-axis is the b^* channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

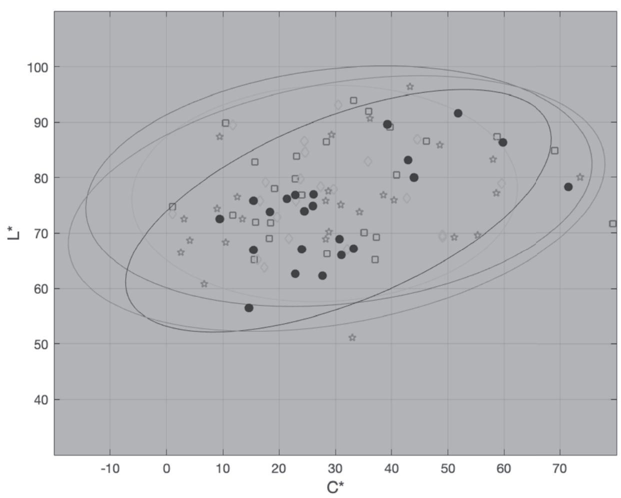

Figure 24. Memory color of beach sand. The x-axis is the chroma, C^* channel, and the y-axis is the lightness, L^* channel. The ellipses define the 90% confidence interval for the data points. The color of ellipses defines its respective color data points.

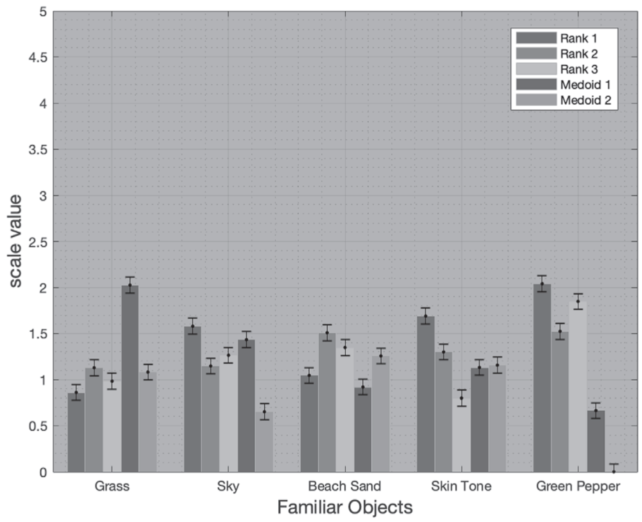

Figure 25. The scale values obtained for five renderings of familiar objects.

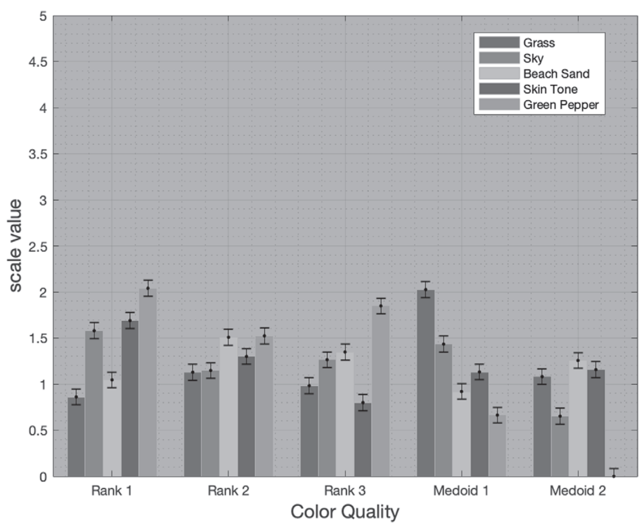

Figure 26. Bar plot for color quality assessment of familiar objects. The x-axis defines the different techniques used to render different color quality images and the y-axis defines the scale value.

observers. For skin, the top preference is the same, but the lowest scaled image is different. For the naïve observers, the lowest scaled image is the rank 3 image, while for the Color Science observers, it is Medoid 1. This difference could be because of the difficult judgment of lower quality. The remaining three images are equally scaled for both sets

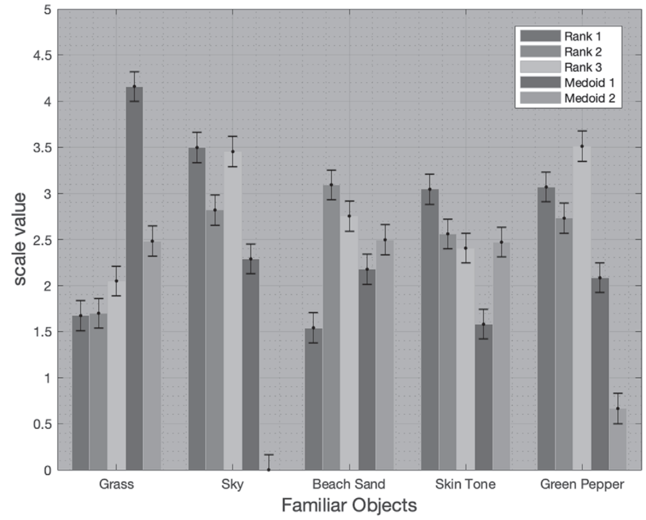

Figure 27. Bar plot for validation of color quality assessment of familiar objects. The x-axis defines the familiar objects and the y-axis defines the scale value.

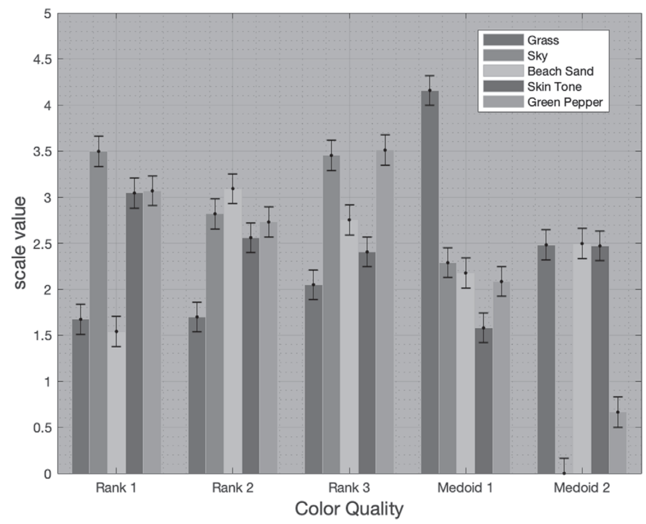

Figure 28. Bar plot for validation of color quality assessment of familiar objects. The x-axis defines the different techniques used to render different color quality images and the y-axis defines the scale value.

of observers. For green pepper, the first and the second preferences have a slight difference, but the preference order for the rest is the same. Overall, these results are in general agreement with the naïve results, thereby providing validation.

In Figure 29, the results suggest that, on average, for both naïve observers and trained color scientists, the familiar image content and textures play an important role in preferred color rendering.

9. CONCLUSION

To improve the color quality for smartphones and digital cameras, it is important to understand user preferences. Color quality assessment is a way to study these preferences. From previous studies, we have seen that there is a correlation

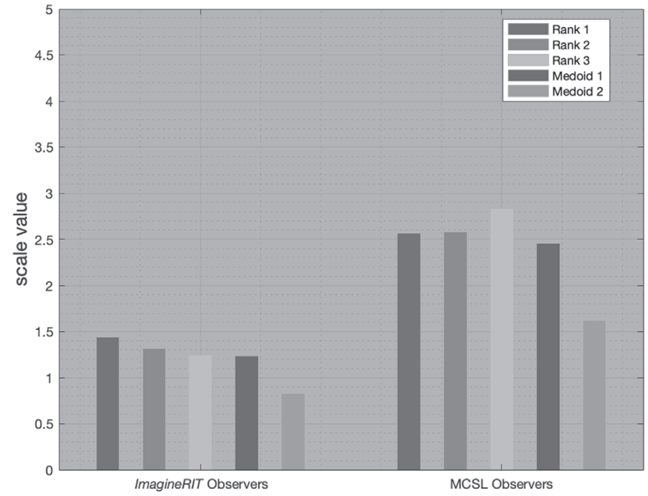

Figure 29. Average bar plot for Imagine RIT observers and the validation by Munsell Color Science Laboratory observers.

between what people like and what people think of when asked to recreate key memory colors. Hence it is also important for the camera industry to study memory color assessment and compare it with the color quality preferences of their users.

This article presents the results of two experiments aimed at better understanding the effect of texture on memory color and perceived color quality. In the first experiment, memory color for familiar scene content with four different types of texture was assessed. Observers were asked to recreate their memory of the color of grass, sky, beach sand, green pepper, and skin by adjusting colors in three dimensions: lightness, redness–greenness, and yellowness–blueness. The four textures were uniform patch, sandstone patch, burlap patch, and image patch as shown in Fig. 7. The results showed that the actual image patch texture type was the most consistent. It is interesting to note that the result for the sky image patch was similar to that for the uniform sky patch texture, and the grass, green pepper skin, human skin, and beach sand images were most similar to Texture 1.

For all of the trials in Experiment 1, except in trials of color adjustment for sky, the uniform image patch had the highest variation. In future studies of memory color adjustment, it may be useful to use a simple texture during the adjustment task in cases where real image content may not be applicable. Additionally, the image patch ellipses were generally within the other three texture ellipses, which indicated that the memory colors were the same despite the inclusion of an artificially imposed texture and not more chromatic as in some studies. More generally, the skin tone ellipses were smaller than those for foliage and sky, which is in strong agreement with previous studies by Bartleson and by Keelan [1, 34]. The main finding of this experiment was that the presence of texture generally decreases the variability in adjustments to memory color.

The second experiment was performed based on the above results using the real object image texture patch and the three top-ranked images from a previous rank order CQA experiment. Color quality assessment was carried out on five different color quality images for each familiar scene element using a paired comparison method. The color quality preference results of this study favored the rank order preference rendered stimuli rather than the memory color rendered stimuli. The results indicated that people prefer a more saturated, airbrushed, and less natural appearance as compared to the memory color. Only for grass, the color quality preference was highest for the memory color rendered image.

10. FUTURE WORK

The results of memory color evaluation and color quality assessment indicated that the hue was generally the most consistent and lightness was also generally consistent. Although saturation was consistent for sky, for the other images, saturation tended to be quite variable. We would like to conduct further study on color-balancing the whole scene. It would be interesting, for example, to evaluate the color quality preference when the hue, chroma, and lightness are changed for sky, grass, and skin with respect to the preferred result for the scenes like those in Fig. 1.

Simple colorimetric accuracy between a scene and its rendering is not sufficient to create an observer preferred image. Additional metrics for describing color rendering performance across subject and viewing conditions—for example, memory colors, particularly skin tones, in mid-day versus sunset, indoor versus outdoor, and fall versus winter—are needed. These factors tend to change the preferences of observers. Another study will be performed on the preference of color quality based on these metrics.

ACKNOWLEDGMENTS

The authors would like to thank Qualcomm for making this study possible as well as all of the observers who participated in the experiments.

REFERENCES

[1] C. J. Bartleson, "Memory colors of familiar objects," JOSA **50**, 73–77 (1960).
[2] S. R. Fernandez, M. D. Fairchild, and K. Braun, "Analysis of observer and cultural variability while generating," J. Imaging Sci. Technol. **49**, 96–104 (2005).
[3] K. A. Smet, "Cross-cultural variation of memory colors of familiar objects," Opt. Express **22**, 32308–32328 (2014).
[4] E. Kirchner, S. G. Kandi, and a. H. Saeedi, "An attempt to reconstruct the meaning of al-Tusi's color words," Color Res. Appl. **41**, 206–216 (2016).
[5] Y. E. A. Zhu, "Long-term memory color investigation: culture effect and experimental setting factors," JOSA **34**, 1757–1768 (2017).
[6] S. Farnand, Y. Jang, C. Han, and H. Hwang, "A methodology for perceptual image quality assessment of smartphone cameras," Electronic Imaging **13**, 1–5 (2016).
[7] A. Anku and S. Farnand, "Color quality and memory color assessment," *IS&T CIC26: Twenty-sixth Color and Imaging Conf.* (IS&T, Springfield, VA, 2018), pp. 116–122.
[8] M. D. Fairchild, *Color Appearance Models* (John Wiley & Sons, Ltd, West Sussex, UK, 2013).
[9] G. K. Adams, "An experimental study of memory color and related phenomena," The Am. J. Psychology **34**, 359–407 (1923).
[10] B. A. C. Newhall, "Comparison of successive with simultaneous color matching," JOSA **47**, 43–56 (1957).
[11] K. L. Kelly, K. S. Gibson, and D. Nickerson, "Tristimulus specification of the Munsell Book of Color from spectrophotometric measurements," JOSA **33**, 355–376 (1943).
[12] J. J. M. Granzier and K. R. Gegenfurtner, "Effects of memory colour on colour constancy for unknown coloured objects," I-Perception 190–215 (2012).
[13] A. C. A. Y. L. Hurlbert, "If it's a banana, it must be yellow: The role of memory colors in color constancy," J. Vision **5**, 787 (2005).
[14] S. N. F. J. J. B. Yendrikhovskij and A. H. D. Ridder, "Representation of memory prototype for an object color," Color Res. Appl. **24**, 393–410 (1999).
[15] J. Pérez-Carpinell, "Familiar objects and memory color," Color Res. Appl. **23**, 416–427 (1998).
[16] B. Marina, W. David, and K. R. Gegenfurtner, "Bias effects of short-and long-term color memory for unique objects," JOSA **33**, 492–500 (2016).
[17] P. Siple and R. M. Springer, "Memory and preference for the colors of objects," Perception & psychophysics **34**, 363–370 (1983).
[18] P. Bodrogi and T. Tarczali, "Colour memory for various sky, skin, and plant colours: Effect of the image context," Color Res. Appl. **26**, 278–289 (2001).
[19] K. A. Smet, "Colour appearance rating of familiar real objects," Color Res. Appl. **36**, 192–200 (2011).
[20] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, "Color appearance of familiar objects: Effects of object shape, texture, and illumination changes," J. Vision **8**, 13 (2008).
[21] M. Vurro, Y. Ling, and A. A. C. Hurlbert, "Memory color of natural familiar objects: Effects of surface texture and 3-D shape," J. Vision **13**, 20 (2013).
[22] E. Hering, *Outlines of a Theory of the Light Sense* (Harvard University Press, 1964).
[23] Y. Ling, "The Colour Perception of Natural Objects: Familiarity, Constancy and Memory," 2006 Doctoral dissertation, Newcastle University (University of Newcastle upon Tyne, 2005).
[24] Y. Yan and H. Suk, "Skin balancing: skin color-based calibration for portrait images to enhance the affective quality," *27th Color and Imaging Conf. Final Program and Proceedings* (IS&T, Springfield, VA, 2019), pp. 91–94.
[25] C. J. Bartleson, "Color in memory in relation to photographic reproduction," Pht. Sci. Eng. **5**, 327–331 (1961).
[26] C. J. Bartleson, "On the preferred reproduction of flesh, blue-sky, and green-grass," Photographic Sci. Engng. **6**, 19–25 (1961).
[27] C. L. Sanders, "Color preferences for natural objects," llum. Engng. **54**, 452–456 (1959).
[28] G. A. Gescheider, *Psychophysics: the Fundamentals*, 3rd ed. (Psychology Press, Lawrence Erlbaum Assoc., NJ, 1997).
[29] E. A. Day, L. Taplin, and R. S. Berns, "Colorimetric characterization of a computer-controlled liquid crystal display," Color Res. Appl. **29**, 365–373 (2004).
[30] E. W. Jin, B. W. Keelan, J. Chen, J. B. Phillips, and Y. Chen, "Softcopy quality ruler method: Implementation and validation," Proc. SPIE **7242**, 724206 (2009).
[31] S. Xue, "Exploring the use of memory colors for image enhancement," Human Vision and Electronic Imaging XIX **9014**, 901411 (2014).
[32] S. P. Farnand and M. D. Fairchild, "Designing pictorial stimuli for perceptual experiments," Appl. Opt. **53**, C72–C82 (2014).
[33] P. G. Engeldrum, "Psychophysical scaling: a toolkit for imaging systems," *Optics & Photonics News* (Imcotekpress, Winchester, MA, 2000).
[34] B. Keelan, *Handbook of Image Quality: Characterization and Prediction* (CRC Press, Marcel Dekker, New York, 2002).