Eye tracking observers during color image evaluation tasks

Jason S. Babcock[1,2*], Jeff B. Pelz[1], and Mark D. Fairchild[2]

[1] Visual Perception Laboratory
[2] Munsell Color Science Laboratory
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, NY, USA 14623

ABSTRACT

Eye movement behavior was investigated for image-quality and chromatic adaptation tasks. The first experiment examined the differences between paired comparison, rank order, and graphical rating tasks, and the second experiment examined the strategies adopted when subjects were asked to select or adjust achromatic regions in images. Results indicate that subjects spent about 4 seconds looking at images in the rank order task, 1.8 seconds per image in the paired comparison task, and 3.5 seconds per image in the graphical rating task. Fixation density maps from the three tasks correlated highly in four of the five images. Eye movements gravitated toward faces and semantic features, and introspective report was not always consistent with fixation density peaks. In adjusting a gray square in an image to appear achromatic, observers spent 95% of their time looking only at the patch. When subjects looked around (less than 5% of the time), they did so early. Foveations were directed to semantic features, not achromatic regions, indicating that people do not seek out near-neutral regions to verify that their patch appears achromatic relative to the scene. Observers also do not scan the image in order to adapt to the average chromaticity of the image. In selecting the most achromatic region in an image, viewers spent 60% of the time scanning the scene. Unlike the achromatic adjustment task, foveations were directed to near-neutral regions, showing behavior similar to a visual search task.

Keywords: Eye tracking, visual perception, looking at pictures, fixation maps, color image quality

1. INTRODUCTION

1.1 Image Quality Judgments and Psychometric Scaling

Experiments focusing on color tolerance for image reproductions (Stokes, 1991; Fernandez, 2002), and the effect of image content on color difference perceptibility (Judd and Wyszecki, 1975; Farnand, 1995) have not investigated how observers look at images during standard psychophysical tasks. Generally, certain assumptions are made regarding the applicability of visual data collected in laboratory experiments. One question is whether the perceptions resulting from psychophysical experiments correlate to visual perceptions in the real world of imaging devices and displays. Further, selecting the best psychophysical technique is often based on the confusion of the sample set, the number of samples used, and observer effort. Practical situations further dictate which method is most fitting. For example, softcopy displays make best use of the paired comparison paradigm over rank order due to the impracticality of displaying many images on the screen while maintaining high-resolution. Assuming all other factors are equal, how well does a scale obtained from one technique compare to that of another? Further, how do we know whether different experimental techniques themselves have any influence on the strategies adopted by observers? Experiment I examines whether viewing strategies are substantially different across: 1) *paired comparison*, 2) *rank order*, and 3) *graphical rating* tasks. Eye movement data collected in this experiment compares the locus of fixation across the three tasks to indicate which regions (in the five images viewed) received the most "foveal" attention, and whether peak areas of attention were the same across the three tasks.

---

[*] Correspondence: babcock@cis.rit.edu

1.2 Color Appearance and Chromatic Adaptation

In attempting to better understand the mechanisms responsible for the stable perception of object color despite changes in illumination and viewing conditions, much research has focused on chromatic adaptation and the effects of simultaneous contrast. Historically, many of these experiments have examined the appearance of uniform color patches presented under conditions where illumination, size, and/or color of the background have been manipulated. More recently, in the context of image reproduction, participants have adjusted patches against variegated backgrounds (Breneman, 1987; Zaidi et al., 1998; Fairchild, 1999; Lee and J. Morovic, 2001) or have manipulated images on a monitor to produce visual matches in cross-media situations (Braun and Fairchild, 1996; Fairchild and Braun, 1997; Fairchild and Johnson, 1999). As these experiments move further away from uniform backgrounds to more spatially and cognitively-complex stimuli such as images, it is important to know whether the history of fixations has any influence on color appearance. It is likely that semantic features in an image, such as faces and memory-color objects, demand more attention during image-quality judgments since observers have an internal expectation (daily experiences and preference) of what these objects should look like (Hunt et al., 1974; Fedorovskaya et al., 1997; Yendrikhovskij et al., 1999). A number of experiments indicate that semantic and informative objects in the scene receive more fixations per observer than other objects in the scene (for a review see Henderson & Hollingworth, 1998). What impact semantic features have on artifact detection and image-quality preference are questions that can be answered by recording where subjects look in an image. Further, it is possible to investigate the history of individual fixations and their impact on the state of chromatic adaptation. Experiment II examines observers' fixation behavior and colorimetric response in: 1) adjusting a patch in an image to appear achromatic, and 2) selecting the most achromatic region in an image.

2. BACKGROUND

2.1 Eye Movements and Visual Perception

People are often surprised when they discover that the central region of the eye occupies the area of highest visual resolution, and that visual acuity in the periphery is quite poor. A closer look reveals that the eye's retina is composed of two types of sensors called rods and cones. Approximately 120 million rod photoreceptors occupy the retina (mostly in the periphery). Approximately 5-6 million cones occupy the central portion of the retina called the fovea. Despite the high sampling density of rods, visual acuity in the periphery is quite poor (Wandell, 1995, pg. 46). The high-resolution cone photoreceptors are packed tightly together near the optical axis. This peak distribution of cones substantially decreases past one degree of visual angle. Unlike rods, each cone photoreceptor in the fovea reports information in a nearly direct path to the visual cortex. In this region of the brain, the fovea occupies a much greater proportion of neural tissue than the rods (Palmer, 1999, pg. 38). Given these characteristics, detailed spatial information from the scene is best acquired through the high-resolution fovea. Our visual system handles bandwidth limitations associated with visual processing by allowing the eyes to shift to areas of interest rapidly and with little effort. It is estimated that a person makes over 150,000 eye movements each day. In the context of natural tasks, eye movements can be described as a combination of fixations and saccades. A *fixation* indicates that the eye has paused on a particular spatial location in the scene. We re-orient the fovea to the next location of interest by rapidly shifting the eye. This movement is called a *saccade*. *Gaze* is the combination of head and eye positioning. Because most of us pay little attention to the fact that vision is sharpest only in the fovea, the active combination of eye and head positioning creates an adequate impression of high-resolution over the entire visual field. Because we use vision as a tool to get information from the scene, point of gaze is closely linked to the stream of attention and perception.

2.2 Eye Movements and Picture Viewing

Buswell (1935) provided the first thorough investigation of eye movements during picture viewing. He showed that observers exhibited two forms of eye-movement behavior. In some cases viewing sequences were characterized by a general survey of the image, where a succession of brief pauses was distributed over the main features of the photograph. In other cases, observers made long fixations over smaller sub-regions of the image. In general, no two observers exhibited exactly the same viewing behavior. However, people were inclined to make quick, global fixations early, transitioning to longer fixations (and smaller saccades) as viewing time increased. A number of experiments since

Buswell have focused on understanding and modeling the role of eye movements in image perception (see Babcock et al. 2002). In general, these experiments have demonstrated that most observers deploy their attention to the same spatial regions in an image, but not necessarily in the same temporal order. They have shown that where people look is not random and that eye movements are not simply bottom-up responses to visual information. Further, these experiments indicate that the level of training, the type of instruction, and observer's background all have some influence on the observer's viewing strategies.

3. METHODS

3.1 Eye Tracking Instrumentation

The Applied Science Laboratory Model 501 eye tracking system was used for all experiments and is described in more detail in Babcock et al., 2001. The main component includes the head mounted optics (HMO), which houses the infrared LED illuminator, a miniature CMOS video camera (sensitive to IR), and a beam splitter (used to align the camera so that it is coaxial with the illumination beam). An external infrared-reflective mirror is positioned in front of the subject's left eye which simultaneously directs the IR source toward the pupil and reflects an image of the eye back to the video camera. A second miniature CMOS camera records the scene from the subject's perspective. This provides a frame of reference to superimpose crosshairs corresponding to the subject's point of gaze. Because the system is based on NTSC video signals, gaze position is calculated at 60 Hz (video field rate). The ASL software allows for variable field averaging to reduce signal noise. Since these experiments were not designed to investigate the low-level dynamics of eye movements, gaze position values were averaged over eight video fields. This yielded an effective temporal resolution of 133 msec. Both horizontal and vertical eye position coordinates with respect to the display plane were recorded using the video-based tracker in conjunction with a Polhemus 3-Space Fastrak magnetic head tracker (MHT). This system uses a fixed transmitter and a receiver attached to the eye tracker headband. Gaze position (integrated eye-in-head and head-position & orientation) is calculated by the ASL using the bright pupil eye image and a head position/orientation signal from the MHT.

To estimate the accuracy of the track across subjects, the average angular distance from known calibration points and fixation record was calculated for both 9 and 17-point targets. Accuracy was examined on data acquired from two displays; a 50" Pioneer Plasma Display (PPD), and a 22" Apple Cinema Display (ACD). On-average, the accuracy of the eye tracker was 1°. However eye movements toward extreme edges of the screen did produce deviations as large as 5°.

3.2 Stimulus Display

Display size is particularly significant in eye movement studies because the accuracy of the track is defined as a function of visual angle. At a constant distance larger displays will result in a smaller fraction of fixation uncertainty within an image. For these reasons a 50 inch Pioneer Plasma Display (PPD) and a 22 inch Apple Cinema Display (ACD) were used to present stimuli for the two experiments discussed in section 3.3 and 3.4. Both displays were characterized using one-dimensional lookup tables followed by a 3x3 matrix as outlined in technical reports by Fairchild and Wyble (1998), and Gibson and Fairchild (2001). Optimal flare terms were estimated using the techniques outlined by Berns, Fernandez and Taplin (in press) and a regression-based channel interdependence matrix was included to further improve the accuracy of the forward models.

3.3 Experiment 1 – Psychometric Scaling Tasks

3.3.1 Experiment Setup

For this experiment the Pioneer Plasma Display and Apple Cinema Display were used for stimulus presentation. Observers performed rank order, paired comparison, and graphical scaling tasks on both displays evaluating the five images in Figure 1. For the plasma display, images were 421 x 321 pixels, subtending 13 x 9° at a viewing distance of 46 inches. For the LCD, images were 450 x 338 pixels with a visual angle of 9.5 x 7° at a distance of 30 inches.


wakeboarder vegetables firefighters kids bug

Figure 1 - Five images used in the psychometric scaling tasks.

For each original shown in Figure 1, five additional images were created by manipulating attributes such as lightness, chroma, or hue. The intention was to simulate the variability from a set of digital cameras or scanners. Adobe Photoshop was used to perform hue rotations for the *kids* and *firefighters* images and chroma manipulations for the *bug* images. The *wakeboarder* and *vegetables* images were manipulated by linearly increasing or decreasing the slope of CIE L^*_{ab} in the original image. Table 1 shows the median pixel-wise color differences from the original image in CIELAB lightness (L^*_{ab}), chroma (C^*_{ab}), and hue (h_{ab}) coordinates for the respective image manipulations using the forward models of the two displays.

Table 1 - Colorimetric manipulations applied to the five images shown in Figure 1

Pioneer Plasma

	wakeboarder	Vegetables	firefighters	kids	bug
manipulations	median ΔL^*_{ab}	median ΔL^*_{ab}	median Δh_{ab}	median Δh_{ab}	median C^*_{ab}
image1	6.60	2.94	original	-9.58	-8.49
image2	original	Original	-5.63	-4.91	-4.17
image3	9.82	4.50	4.91	original	original
image4	15.97	6.05	9.53	4.63	4.45
image5	-3.13	-2.60	13.59	8.96	10.58
image6	-9.01	-1.32	17.66	13.18	18.92

Apple Cinema

	wakeboarder	Vegetables	firefighters	kids	bug
manipulations	median ΔL^*_{ab}	median ΔL^*	median Δh_{ab}	median Δh_{ab}	median C^*_{ab}
image1	6.05	2.46	original	-8.87	-10.06
image2	original	Original	-5.20	-4.56	-4.95
image3	9.18	3.74	4.46	original	original
image4	15.35	5.01	9.02	4.28	5.39
image5	-2.81	-1.06	13.01	8.45	12.97
image6	-8.06	-2.06	16.90	12.44	23.58

Nineteen subjects, (5 females, 14 males,) ranging from 19-51 years of age participated in this experiment. Eye tracking records from six of the subjects were discarded due to poor calibration, excessive number of track losses, and problems related to equipment failure. Psychophysical data was collected and analyzed for all 19 observers. Stimulus presentation for the rank order, paired comparison and graphical rating tasks was implemented as a graphical user interface in Matlab. The rank order task (Figure 2a) displayed all six manipulations and observers ranked them from 1 to 6, where 1 was most preferred and 6 was least preferred. In the paired comparison task (Figure 2b) subjects used the mouse to select the most preferred image of the two images displayed on screen. All pairs of the 6 manipulations were displayed in random order. In the graphical rating task (Figure 2c) subjects used a slider bar to rate the image quality between two imaginary extremes (no anchor pairs were given).

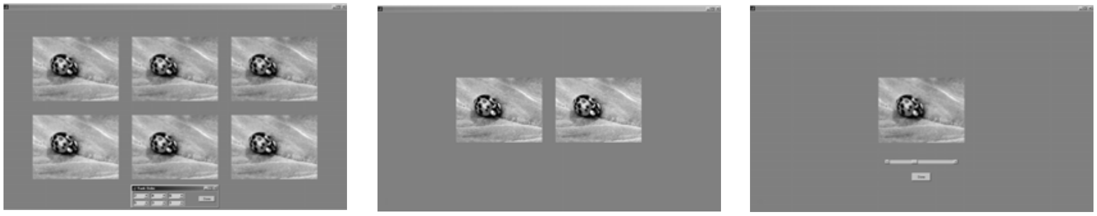
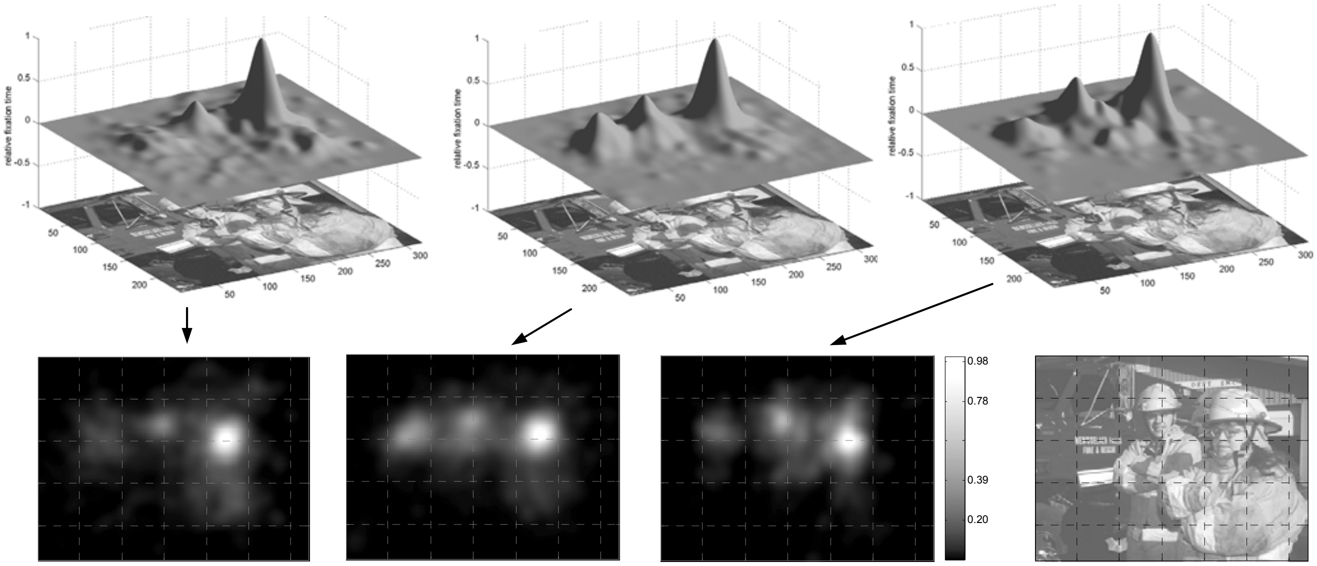

a) rank order b) paired comparison c) graphical rating

Figure 2 – Layout of the rank order, paired comparison, and graphical rating experiments.

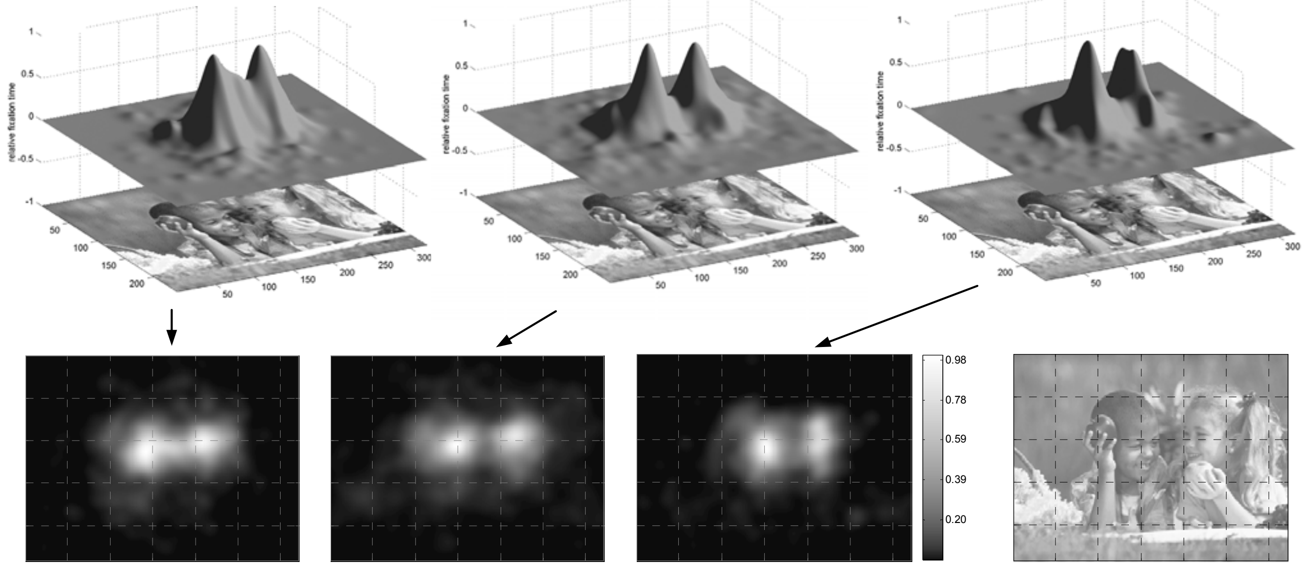
3.3.2 Fixation Duration Results

For 13 subjects, five image groups, and two displays, fixation duration showed that viewers spent about 4 seconds per image in the rank order task, 1.8 seconds per image in the paired comparison task, and 3.5 seconds per image in the graphical rating task. Although the amount of time subjects spent looking at images for each of the three tasks was different, peak areas of attention, as indicated by fixation density maps, show a high degree of similarity. Clearly certain objects in the scene received more fixation attention than other objects (e.g. the right firefighter's face in Figure 3, and the faces in Figure 4).



a) rank order b) paired comparison c) graphical rating

Figure 3 – The top graphs show normalized fixation density across 13 subjects for the rank order, paired comparison, and graphical rating tasks for the *firefighters* image (viewed on a 50" plasma display). The bottom graphs show the same data as a gray scale image.



a) rank order b) paired comparison c) graphical rating

Figure 4 – The top graphs show normalized fixation density across 13 subjects for the rank order, paired comparison, and graphical rating tasks for the *kids* image (viewed on a 50" plasma display). The bottom graphs show the same data as a gray scale image.

The fixation density maps across the rank order, paired comparison, and graphical rating tasks appear similar. To quantify this similarity the fixation maps were treated as images (shown in the bottom graphs of Figures 3 & 4) and the 2-D correlation between task pairs was computed using Equation 1. The 2-D correlation metric is sensitive to position and rotational shifts and provides a first-order measure of similarity between two grayscale images (Russ, 1994; Gonzalez & Woods, 2001). Table 1 presents the correlations calculated between fixation maps for all pairs of the three scaling tasks.

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \overline{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \overline{B})^2\right)}} \qquad (1)$$

Table 1 Correlation (r) between rank order, paired comparison, and graphical rating tasks for two displays

Pioneer Plasma	*wakeboarder*	*vegetables*	*firefighters*	*kids*	*bug*
2-D correlation between:					
Rank order & Paired comp	0.90	0.74	0.90	0.94	0.93
Rank order & Graphical rating	0.86	0.62	0.92	0.93	0.92
Paired comp & Graphical rating	0.93	0.82	0.87	0.92	0.90

Apple Cinema	*wakeboarder*	*vegetables*	*firefighters*	*kids*	*bug*
2-D correlation between:					
Rank order & Paired comp	0.91	0.60	0.96	0.96	0.95
Rank order & Graphical rating	0.89	0.56	0.90	0.94	0.93
Paired comp & Graphical rating	0.94	0.86	0.89	0.94	0.95

Table 1 shows that the *vegetables* image produced the lowest overall correlation between the three tasks, and that rank order fixation maps compared to graphical rating fixation maps were most different. This result is likely attributed to the spatial complexity of the image and the variety of objects with distinct memory colors. Highlight regions on the mushrooms and cauliflower objects were clipped for boosts in lightness. These objects seemed to attract a high degree of attention, but not with the same weight. Because the vegetables scene had over 20 distinctly named objects, it is also likely that observers moved their eyes toward different regions out of curiosity, causing unique fixation maps across tasks.

The *bug* and *kids* images resulted in the highest overall correlations across tasks. This result is likely related to the fact that semantic features were located mostly in the center of the image and that surrounding regions were nearly uniform with low spatial frequency and moderate color changes. Since flesh tones are important to image quality judgments (Hunt et. al., 1974), fixation duration was expected to be high for faces in the *wakeboarder*, *firefighters*, and *kids* images.

3.3.3 Circling Regions Used to Make Preference Decisions

There is some debate as to whether regions of interest could just as easily be obtained by having viewers physically mark or circle important regions in the image. One question is whether regions with a higher number of fixations correspond to regions identified by introspection. To make this comparison, subjects were given a print-out (at the end of the experiment) showing the five images in Figure 1. Directions on the sheet instructed observers to: "Please circle the regions in the image you used to make your preference decisions." Each participant's response was reconstructed as a grayscale image in Adobe Photoshop. Circled regions were assigned a value of 1 digital count and non-circled areas were assigned a value of 0 digital counts. Figure 5a shows an example across 13 subjects for the *firefighters* and *kids* images. Figure 5b shows fixation density and the regions of importance circled for a single subject. In some cases, subjects circled areas that received very few fixations. The leftmost image in Figure 5b shows a subject's eye movement record collapsed across all observations of the bug for both displays. The areas he circled are superimposed over the image indicating that the bottom portion of the leaf was important to his preference decisions. However, very few fixations occurred in those regions. Inconsistencies were also evident in eye movement records for three other subjects looking at the *firefighters* and *kids* images (shown in Figure 5b). It is evident that subjects' peak areas of attention do not necessarily agree with introspective report.

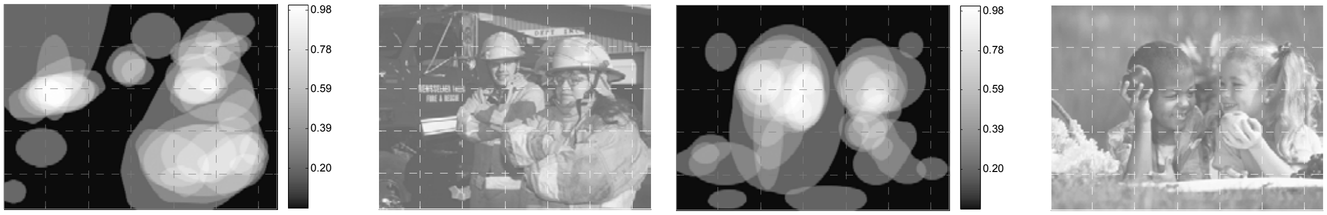


Figure 5a –Subjects circled the regions in the image they used to make their preference decisions. Plots are normalized to the region with the highest sum across grayscale images from 13 observers.

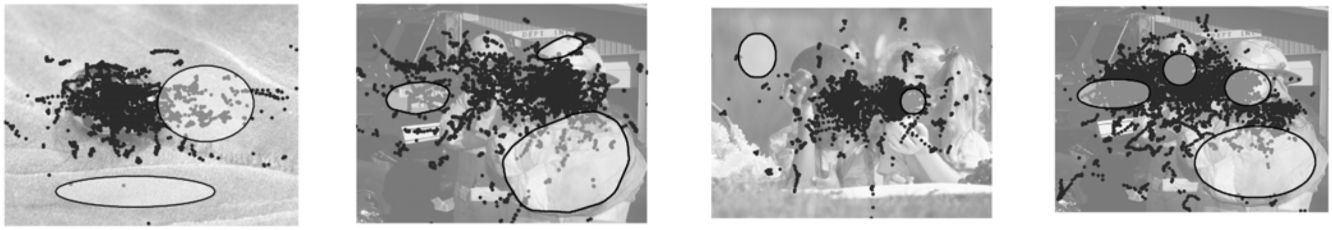


Figure 5b —Black markers indicate fixations compiled across the six manipulations for both displays from a single individual. Circles indicate regions in the image that were important to the observer's preference decision.

3.4 Experiment 2 – Achromatic Patch Adjustment and Selection

The first task in Experiment 2 examined how the white point, spatial complexity, and semantic features in an image influenced observers' viewing behavior when asked to make a patch in the center of an image appear achromatic. One hypothesis is that subjects look to areas in the image that are near neutral to ensure that their patch adjustment appears achromatic in the context of the scene. This hypothesis suggests that features such as shadows and other gray objects will serve as a frame of reference in determining what is neutral. If this is true, observers should actively seek out gray features in the image. Another question to be answered is whether local adaptation from previous fixations causes subjects' color patch adjustments to be skewed toward the mean chromaticity of recently fixated objects.

3.4.1 Task 1 – Patch Adjustment

Seventy-two images (640 x 410 pixels), randomized for each observation, were viewed on a 50" Pioneer Plasma Display. Images subtended 27 x 17° (slightly larger than an 11x 17" page from a distance of 46 inches), and the remaining area on the screen was set to zero digital counts. Thirty-six images had the default white point of the monitor whose correlated color temperature approximated CIE illuminant D65 (6674 K). The other 36 images were manipulated to have a white point that approximated D93. Both the D65 and D93 image groups were split into three categories: 1) the original photograph (labeled as N for normal), 2) a mosaic version of the original (labeled as M for mosaic), and 3) a spatially uniform gray (G for gray) whose digital counts were the mean tristimulus values of the N and M images.

a) Normal (N) b) Mosaic (M) c) Gray average (G)

Figure 6 – Example images used in Experiment II, task 1. Subjects manipulated the central gray square (subtending 2° visual angle) using the four arrow keys.

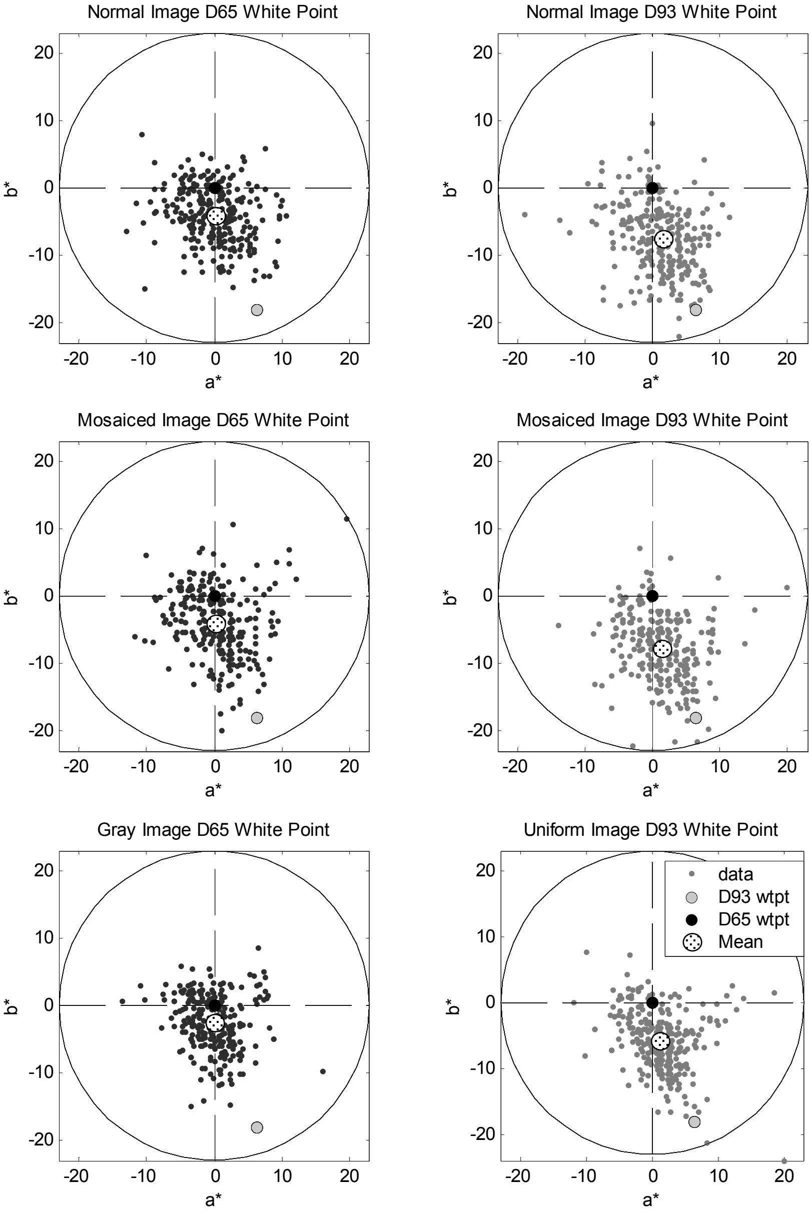
The color appearance of the central 2° patch was controlled in CIELAB color space using the monitor's white point as the reference white. At the start of each presentation, the center patch was set to a predefined color ranging from ± (5 to 10) a* b* units. Pressing one of the four arrow keys changed the patch 0.75 units in the selected opponent direction. The lightness of the patch remained constant throughout the experiment (L* equaled 60). Participants marked completion of their final adjustment by hitting the return key. This event signaled the program to advance to the next trial and to store the history of colorimetric adjustments as RGB, XYZ, and CIELAB coordinates. Between each trial a neutral background (with L* = 50) appeared for 15 seconds. Subjects were instructed to fixate on a series of count-down numbers as they appeared randomly in one of ten locations on the screen. The 15 second pause was used to re-adapt the subject to the monitor's D65 white point and to cancel out any afterimages resulting from the previous trial.

3.4.2 Adaptation Results

As might be expected from chromatic adaptation, there was a -3.60 b* shift toward the blue in the mean D93 color adjustments. To test whether D65 and D93 means were statistically different, a MANOVA was performed using the a* b* coordinates as the response variables and an index (1 for D65, and 2 for D93) as the model.

Fairchild and Lennie (1992), and Fairchild and Reniff (1995) expressed percentage of adaptation as the percentage of the Euclidian distance from the D65 u'v' coordinate to the adapting u' v' coordinate. Observers' patch adjustment for images with a D93 white point indicate about 65% adaptation. Given that the average patch adjustment time was 25 seconds, 65% adaptation agrees well with the time course of adaptation suggested in the Fairchild and Reniff study.

Figure 7 examines the colorimetric results by separating the data to N, M, and G groups to see whether patch adjustment were different across these categories. Row plots denote N, M, G; columns indicate D65 on the left and D93 on the right. MANOVA t-tests between N-G, and M-G, show that color adjustments between normal and mosaic images are statistically different from the gray averaged images (p-value < 0.001). However, comparing normal images with mosaic images produced no statistical difference (p-value = 0.56).


Normal Image D65 White Point

Normal Image D93 White Point

Mosaiced Image D65 White Point

Mosaiced Image D93 White Point

Gray Image D65 White Point

Uniform Image D93 White Point

Figure 7 – Plots subject's final patch adjustments for N, M, and G images groups. The black marker represents the D65 a* b* white point and the white marker represents the D93 white point. The dotted marker represents the mean a* b* for the data in each plot.

3.4.3 Percentage of Surround Fixations

This section examines the amount of time that was spent fixating on the patch region compared to the time spent fixating on the surrounding image. One hypothesis is that subjects make more exploratory eye movements in the normal (N) photograph than in the mosaic (M) or gray averaged (G) images because semantic features in the scene tend to elicit viewers' interest. To examine this hypothesis, 2-D fixation histograms were generated for each subject across all images. Fixations falling inside a 50 pixel radius from the center (where the patch was) were defined as patch fixations and fixations falling outside this region were defined as surround fixations (see Figure 8). The percentage of surround fixations was computed using Equation 2.

$$(2) \qquad SF = \frac{s}{N}100$$

SF = % surround fixations
s = number of fixations occuring outside the 50 pixel radius
N = total number of fixations on the whole image (patch + surround)

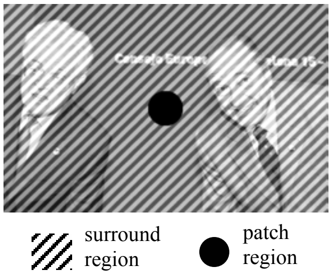


Figure 8 – Example images used in Experiment II, task 1. Subjects manipulated the gray square (subtending 2° visual angle) using the four arrow keys.

The percentage of surround fixations between D65 and D93 images was not statistically different at a 95% confidence level (p-value = 0.125). Paired t-tests between means for the N, M, and G categories indicate that viewing behavior is statistically different when comparing the fixations from the normal (N) and mosaic (M) images to the gray averaged (G) images (in both cases p-value < 0.001). The percentage of fixations on the surround was about twice as much for the N and M images in comparison to the gray-average images. Altogether, less than 5% of the viewing time was allocated to the surround, regardless of N, M, or G types. This low percentage was not expected for the normal (N) image, and illustrates how task dependencies influence eye movement behavior.

Examination of the video record and colorimetric plots of fixation history (detailed in Babcock, 2002) indicate that viewers do not deliberately seek out near-neutral objects to ensure that their patch adjustment appear achromatic in the context of the image. This suggests that people have a strong impression of gray and do not rely on features in the scene to validate their judgment of gray. Furthermore, less than 5% of the total patch adjustment time was spent looking around the image. These fixations occurred early during the trial and were consistently directed toward people and faces, not shadows or achromatic regions.

3.4.4 Task 2 – Selecting the Most Achromatic Region

Task 2 examined eye movement behavior when the subject's goal was to select the most achromatic region in an image. The layout for the achromatic selection interface was similar to Task 1 with the exception that the gray patch in the center of the screen was removed. Forty-eight of the images used in Task 1 were used for Task 2. The image set was randomized for each subject and consisted of D65-D93 normal (N) and mosaic (M) categories. This excluded all gray averaged (G) images. Observers were instructed to use the mouse to select the region in each image that appeared the most achromatic.

3.4.5 Percentage of Surround Fixations

As in section 3.4.3, fixation histograms were generated for each subject across all images. Fixations falling inside a 50 pixel radius from the observer's mouse click were defined as target fixations and fixations falling outside this region were defined as surround fixations. The percentage of fixations on the surround for each subject was computed using Equation 2.

The fraction of surround fixations between D65 and D93 images was not statistically different at a 95% confidence level. Paired t-tests between means for the N, and M categories indicate that viewing behavior was not statistically different (p-value = 0.173) when comparing the fixations from the normal (N) and mosaic (M) images. Roughly 60% of the viewing time was allocated to the surround and 40% to the target area. Figure 9 plots the frequency of fixations on the surround across N and M images for both the patch adjustment task (Task 1) and achromatic selection task (Task 2). This provides a clear example of how task dependencies can influence eye movement behavior.

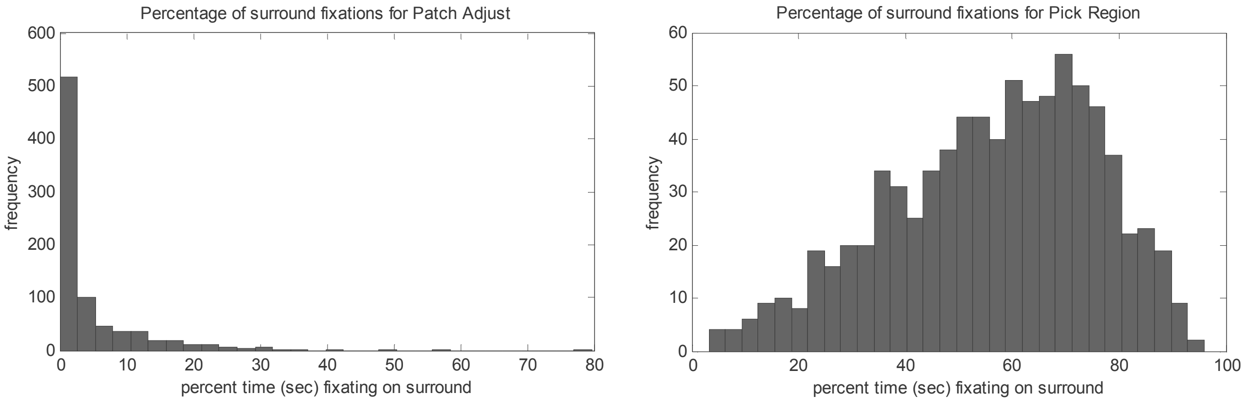


Figure 9 – The left graph plots mean % fixation on the surround for N and M images from the patch adjustment task (Task 1). The right graph plots mean % fixation for N and M images from the achromatic patch selection task (Task 2).

4. DISCUSSION AND CONCLUSIONS

4.1 Eye Movements and Psychometric Scaling

4.1.1 Rank Order, Paired Comparison, and Graphical Rating Tasks

Fixation duration plots from the rank order task showed that people spend roughly the same amount of time looking at each of the six manipulations, but different amounts of time per image type. Video records indicate that observers typically rank the highest and lowest images first, making several fixations to these "reference" images while finalizing ranks among the remaining images. In the paired comparison task there was no tendency to fixate longer on the left or right image, however, subjects did spend more time looking at images that were preferred versus images that were not preferred (0.28 seconds more time for preferred images). Video records indicate that judgments were performed quickly; usually making from 2 to 4 saccades between images before advancing to the next pair. Unlike the other scaling tasks, the graphical rating task resulted in very different fixation behaviors across the five image types. For images with lightness manipulations (*wakebaorder* and *vegetables* images), observers spent more time looking at images rated higher on the preference scale than images rated lower on the preference scale. However, for the chroma manipulation (*bug* image) and one of the hue manipulations (*kids* image), more time was spent looking at images falling in the middle of the preference scale. This behavior was consistent across both displays, and indicates that observers thought carefully about where particular images belonged on the preference continuum.

4.1.2 Peak Areas of Attention

The spatial distribution of fixations across rank order, paired comparison, and graphical rating tasks showed a high degree of consistency. Observers' peak areas of attention gravitated toward faces and semantic regions as reported in many eye tracking studies (see Henderson & Hollingworth, 1998). However, the vegetables scene, which contained over 20 identifiable objects, generated the lowest correlation between the three tasks. It is hypothesized that the spatial complexity, high number of objects with memory colors, and/or observer curiosity may have caused different viewing behaviors across the three tasks.

4.1.5 Introspection and Scaling

Section 3.3.3 showed that introspective report, as indicated by circling regions in the image at the end of the experiment, was not always consistent with image foveations. Furthermore, the spatial weighting implied by introspection maps is broader than is implied by eye movement maps. Psychophysical results across rank order, paired comparison, and graphical rating tasks generated similar, but not identical, scales values for the *firefighters*, *kids*, and *bug* images. Given the similarity between fixation densities across the three tasks, the differences in scales are probably related to statistical treatment and image confusability, rather than eye movement behavior. However, the small number of subjects (19 in this case) and unanimous agreement across paired comparison and rank order judgments will require a larger number of observers to validate scale similarity across the three tasks. The implications of scale similarity are important because it means that scale values obtained from one type of experiment can be directly compared to scale values from another type of experiment.

4.2 Achromatic Patch Adjustment and Selection

Experiment 2 examined observers' visual strategies when asked to perform achromatic patch adjustments in scenes that varied the spatial complexity and semantic content. These results were compared with a second task that had observers select the most achromatic region from the same set of images.

4.2.1 Achromatic Patch Adjustment

More than 95% of the total patch adjustment time was spent looking strictly at the patch. This result shows that even when participants are allowed to move their eyes freely, putting an adjustment patch in the center of the screen discourages people from viewing the image in a natural way. When subjects did look around (less than 5% of the time), they did so early during the trial. These foveations were consistently directed toward people and faces, not shadows or achromatic regions. This result shows that viewers do not deliberately seek out near-neutral objects to ensure that their patch adjustments appear achromatic in the context of the scene. They also do not scan the image in order to adapt to a gray world average. Apparently people have a strong internal representation of gray, and do not rely on features in the scene to validate their patch adjustment (i.e. their "definition" of gray).

The percentage of exploratory fixations in the image (the 5% surround fixations) was statistically different between normal images (N), mosaic images (M), and uniform gray-averaged images (G). Differences were highest between normal vs. gray-averaged (N-G) and mosaic vs. gray-averaged (M-G) pairs. This result indicates that observers do not look around as much in surrounds with a gray-average. This behavior may be responsible for tighter variances in color adjustment data for the G images as compared to the N and M images.

As demonstrated in other studies, the mean chromaticity of the image influenced observers' patch adjustments. Adaptation to the D93 white point was about 65% complete from D65. This result agrees reasonably with the time course of adaptation occurring over a 20 to 30 second exposure to the adapting illuminant, which was about the mean time spent performing each adjustment trial (Fairchild and Reniff, 1995). Images whose mean a* b* coordinates were near-neutral also resulted in adjustments falling along the D65-D93 white-point line. Fixations to faces and semantic features in the scene did not appear to alter observers' achromatic adjustments. It was difficult address the history of fixations on adaptation further since only 5% of observers' fixations were allocated to areas other than the patch.

4.2.2 Achromatic Patch Selection

Viewers spent 60% of the time scanning the scene in order to select the most achromatic region in the image. Unlike the achromatic patch adjustment task, subjects' foveations were consistently directed toward achromatic regions and near-neutral objects. Eye movement records showed behavior similar to what is expected in a visual search task. The percentage of surround fixations between N and M categories were not statistically different.

4.2.3 Recommendations

Because it was difficult address the history of fixations on adaptation (since subjects spent so little time looking around in the image), a future revision of this experiment might have the observer free-view an image, and then display several near-neutral patches. The observer's task would be to select the most achromatic patch as quickly as possible.

This task would elicit more realistic viewing behavior and would allow for a more interesting history of fixations. This experiment could be further expanded by comparing eye movement behavior in real scenes verses soft-copy image displays.

4. ACKNOWLEDGMENTS

This work was supported in part by the Naval Research Laboratory, the New York State Office of Science, Technology, and Academic Research, and the Eastman Kodak Company. This work could not have been accomplished without the assistance of students, faculty, and staff at the Chester F. Carlson Center for Imaging Science. Thanks for being subjects in my experiments!

5. REFERENCES

Babcock, J. S. (2002). *Eye Tracking Observers During Color Image Evaluation Tasks*. Master's Thesis. New York: Rochester Institute of Technology.

Babcock, J.S., Lipps, M., Pelz, J.B. (2002). How people look at pictures before, during and after scene capture: Buswell revisited. In B.E.Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging V, SPIE Proceedings*, **4662**, 34-47.

Berns, R. S., Fernandez , S.R., Taplin, L., (2002). Estimating black level emissions of computer-controlled displays. *Color Res. Appl*., (in press).

Braun, K.M. & Fairchild, M.D. (1996). Psychophysical generation of matching images for cross-media color reproduction. *IS&T/SID Color Imaging Conference*, **4**, 212-220.

Breneman, E. (1987). Corresponding chromaticities for different states of adaptation to complex visual fields. *J. Opt. Soc. Am. A,* **4**(6), 1115-1129.

Buswell, G.T. (1935). *How People Look at Pictures*, Chicago: Univ. Chicago Press.

Fairchild, M. D. & Lennie, P. (1992). Chromatic Adaptation to Natural and Incandescent Illuminants, *Vision Res.* **32,** No. 11, 2077-2085.

Fairchild, M. D. & Reniff, L. (1995). Time course of chromatic adaptation for color-appearance judgments, *J. Opt. Soc. Am. A,* **12,** 824-833.

Fairchild, M.D. & Braun, K.M. (1997). Investigation of color appearance using the psychophysical method of adjustment and complex pictorial stimuli. *AIC Color*, 179-186.

Fairchild, M.D., & Wyble, D.R. (1998). Colorimetric Characterization of the Apple Studio Display (Flat Panel LCD). *Munsell Color Science Laboratory Technical Report*, July.

Fairchild, M.D. (1999). A Victory for Equivalent Background – On Average. *IS&T/SID Color Imaging Conference.* **7**, 87-92

Fairchild, M.D. & Johnson, G.M. (1999). Color-appearance reproduction: visual data and predictive modeling. *Color Res. Appl.,* **24,** 121-131.

Farnand, S.P. (1995). *The Effect of Image Content on Color Difference Perceptibility*. Master's Thesis. New York: Rochester Institute of Technology.

Fedorovskaya, E. A., de Ridder, H., Blommaert, F.J.J. (1997). Chroma Variations and Perceived Quality of Color Images of Natural Scenes. *Color Res. Appl*., **22,** 96-110.

Fernandez, S. R. (2002). *Preferences and Tolerances in Color Image Reproduction*. Master's Thesis. New York: Rochester Institute of Technology.

Gibson, J. E. & Fairchild, M.D. (2001). Colorimetric Characterization of Three Computer Displays (LCD and CRT), *Munsell Color Science Laboratory Technical Report*, January.

Henderson, J.M. & Hollingworth, A. (1998). Eye movements during scene viewing: an overview. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 269-293). New York: Elsevier.

Hunt, R.W., Pitt, I.T., Winter, L.M. (1974). The Reproduction of Blue Sky, Green Grass and Caucasian Skin in Color Photography. *J. Photogr. Sci*., **22,** 144-150.

Judd, D.B. & Wyszecki G. (1975). *Color In Business Science, and Industry*. New York: Wiley.

Lee, S.M.M. & Morovic, J. (2001). What Do Complex Backgrounds Integrate To? *IS&T/SID PICS Conference Proceedings*. 15-18.

Palmer, S.E. (1999). *Vision Science Photons to Phenomenology*. Cambridge, MA: MIT Press.

Stokes, M. (1991). *Colorimetric Tolerances of Digital Images*. Master's Thesis. New York: Rochester Institute of Technology.

Wandell, B.A. (1995). *Foundations of Vision*. Sunderland, MA: Sinauer.

Yendrikhovskij, S.N., Blommaert, F.J.J., de Ridder, H. (1999). Color Reproduction and the Naturalness Constraint. *Color Res. Appl*., **24,** 52-67.

Zaidi, Q., Spehar, B., DeBonet, J. (1998). Adaptation to textured chromatic fields. *J. Opt. Soc. Am. A,* **15**(1), 23-32.