

# Chapter 7

## 7. Conclusions and Recommendations

Overall, the objectives of this research project were met. The first goal was to connect what we know about eye movement research with studies regarding image quality evaluation and chromatic adaptation. The second goal focused on learning where people center their attention during color preference judgments, examining the differences between paired comparison, rank order, and graphical rating tasks, and determining what strategies are adopted when selecting or adjusting achromatic regions on a soft-copy display. The third goal was to develop a software library in Matlab to aid in data collection, analysis, and visualization. This library now includes routines for blink removal, saccade interval extraction, offset correction, visualization of fixation density, and GUIs for rank order, graphical rating, and paired comparison scaling experiments. These tools have provided a framework for integrating eye tracking research with image quality studies.

## **7.1 Eye Movements and Psychometric Scaling**

Chapter 5 investigated visual behavior in the context of image quality evaluation. 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.

**7.1.1 Rank Order** - 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.

**7.1.2 Paired Comparison** - 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.

**7.1.3 Graphical Rating** – 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.

**7.1.4 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 (Buswell, 1935; Brandt, 1945; Yarbus, 1967; 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.

**7.1.5 Introspection and Scaling** - Chapter 5 also showed that introspective report, as indicated by circling regions in the image at the end of the experiment, was not always consistent with where people foveated. 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.

**7.1.6 Recommendations** – The most obvious direction for future work is to expand this research to include more observers, more images, additional psychometric scaling techniques, and a larger range of spatial and colorimetric manipulations. While visual behavior was quite similar across the three tasks, extended research will clarify what differences in visual behavior arise for scenes containing a large number of objects (i.e. like the *vegetables* scene).

In developing automatic saliency detectors, it is clear that face/person detection is one of the first steps toward mimicking where people look in scenes. Eye movement maps have may prove valuable to researchers developing image difference and image quality models. With this in mind, a future goal is to develop an on-line eye movement database for people who do not have time or access to eye tracking equipment, but are interested in knowing where people look for a specific set of images.

Experiments in chapter 5 examined eye movement behavior for soft-copy displays in a controlled laboratory setting. In actuality, people are really faced with image quality decisions when reading magazines, watching television, shopping in stores, or looking at posters. One of the next steps might be to examine whether peak areas of attention change when subjects perform hard-copy image quality experiments. Further, it might be interesting to set-up a less formal study, implemented under more realistic situations, such as ranking the image quality of posters in a busy hallway, or rating the capture quality of digital cameras. Babcock *et. al.* (2002) have already conducted portable eye tracking studies aimed at understanding how people look at digital images before, during,

and after scene capture. It seems reasonable to expand this type of experiment to include graphical rating or rank order tasks as well.

## **7.2 Achromatic Patch Adjustment and Selection**

Chapter 6 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.

**7.2.1 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 freely move their eyes, 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.

**7.2.2 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 as would be expected. Eye movement records show 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.

**7.2.3 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.