



Identifying regions of interest in reading an image



Cheng-Min Tsai^{a,b,*}, Shing-Sheng Guan^a

^a Graduate School of Design, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan, ROC

^b Department of Creative Product Design and Management, Far East University, Tainan 74448, Taiwan, ROC

ARTICLE INFO

Article history:

Received 16 August 2014

Received in revised form 19 July 2015

Accepted 2 August 2015

Available online 3 August 2015

Keywords:

Region of interest

Image quality assessment

Eye movement

Eye-tracking device

ABSTRACT

The aim of this study is to develop an effective method to analyze regions of interest (ROIs). Two experiments were conducted at different times using different groups of observers with different images on different displays. Observers' eye-movement data were collected. Fixation maps showing CIELAB L^* values were created. The ΔL^* values between the two maps were used to quantify differences in visual fields, counting methods, observer variability and repeatability between the two experiments.

The results showed that fixation maps can be used to effectively analyze the distribution of eye movements between images. The ΔL^* value calculated for two fixation maps is easy to understand and computes differences based only on ROIs more effectively than differences based on the entire image. The results from the two experiments were consistent, indicating that eye-tracking data are robust for evaluating image quality.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

One of the goals of imaging research is to develop a metric based on image statistics to assess images. Earlier metrics, such as the peak signal-to-noise ratio (PSNR) [1] and root mean square error (RMSE) [2], were focused on the physical measurement of image quality and do not correspond well with the results of visual assessment [3–5]. More recent studies were conducted based on visual assessments using psychophysical methods to investigate parameters that affect the judgment of image quality, such as the naturalness, colorfulness, and sharpness of the images. Efforts were then made to develop color models that predict these parameters and pool the individual parameters to form an overall image quality index [6–13].

The above methods were based on a global analysis of the entire image. Another approach is to identify the regions of interest (ROIs) in an image, defined as the areas of an image that attract more visual attention than the other areas [14,15]. Privitera and Stark have noted that ROIs are defined as the loci of the human's eye fixations and that they can be analyzed using their spatial distribution over the visual stimulus and their temporal ordering [14]. Privitera and Stark have developed an image quality model based on the focus of visual attention within an image rather than the entire image [15,16]. Privitera et al. [15] have conducted a series

of experiments to fit the measured scan path data and identify ROIs, first using an eye-tracking system. They then used geometrical spatial kernels and linear filter models to locate the ROIs in an image. Nguyen et al. [3] have grouped ROIs based on an analysis of scan-paths and sequences of fixation for viewing grayscale images and subsequently performed compression based on the ROIs of an image. The algorithm only addresses grayscale images and thus may not work well for color images.

Because the eye is the first element of the visual system to receive visual information, it is also the only means by which the brain obtains external images. Henderson and Hollingworth [17] have found that eye movements are critical for the efficient and timely acquisition of visual information during complex visual-cognitive tasks. The eye-tracking technique has been widely applied in various research areas, such as human factor and interface design, advertising and marketing, psychology and neuroscience, attention span studies and visual text analysis. In the image assessment research field, fixation-map analysis provides an opportunity to objectively define the principal ROIs for viewing images [18]. This study differs from the previous studies in that it employs an eye-tracking technique to develop an effective method to analyze ROIs.

2. Experimental design

Two experiments for assessing image quality were conducted to test the reliability of the results from the eye-tracking experiment. These experiments were conducted at different times using

* Corresponding author at: Graduate School of Design, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan, ROC.

E-mail address: ansel.tsai@gmail.com (C.-M. Tsai).

different groups of observers with different images on different displays that have different sizes and peak white values. The results from the two experiments can be compared to test the robustness of the method. Table 1 summarizes the experimental conditions in these two experiments.

2.1. Eye-tracking technique

A device used for measuring eye movements is referred to as either an eye-tracking device or an eye tracker [19]. An EyeLink II (SR Research Ltd., Mississauga, Ontario, Canada) with an infra-red head-mounted, video-based, pupil and corneal reflection eye-tracking apparatus was used in this study. This device has high resolution (noise-limited at $<0.01^\circ$) and a fast data-acquisition rate (more than 250 samples per second). It was easy to operate and required less than a minute to conduct the calibration process before commencing the experiment. Fig. 1 shows the experimental conditions. Observers were seated in front of an LCD-TV and wore a headset containing a camera to record their eye movements and fixation locations (see Fig. 2). The time that users fixated on each pixel and the eye positions for each image were stored on the Host PC (personal computer). The observers judged the image quality on a stimulus monitor controlled by the Display PC. The eye-movement data were processed instantly to yield the fixation position and fixation duration in a form ready for use in the data analysis.

2.2. Experiment I

2.2.1. Display

A 30-inch LCD TV was used to display images. A Gretag Macbeth Eye-One colorimeter was used to establish the ICC profile for display characterization. The study was conducted in a laboratory with an ambient illuminance of approximately 230 lux and a correlated color temperature (CCT) approaching 6500 K.

2.2.2. Observers

Thirty observers participated in Experiment I (19 females and 11 males whose average age was 23). All were staff members and postgraduate students from the School of Design at the National Yunlin University of Science and Technology. All had normal color vision according to the Ishihara color vision test.

Table 1
Conditions for Experiments I and II.

Materials	Experiment I	Experiment II
Observers	30	15
Images	11	6
Display	Sharp 30" LCD-TV	SONY 40" LCD-TV
Peak white (x, y chromaticity coordinates)	(0.3127, 0.3285)	(0.3153, 0.3297)
Peak white (luminance (cd/m ²))	140	203
Mid-gray background color	$L^* = 61.96, a^* = -1.21, b^* = -13.21$	$L^* = 74.75, a^* = 1.45, b^* = -12.84$
Resolution of the display	1280 × 768 pixels	1920 × 1080 pixels
Image size (pixels)	1280 (W) × 768 (H)	768 (W) × 512 (H)
Viewing distance	120 cm	70 cm
Viewing angle (degree)	$\approx 30.56^\circ$ (W) × 18.38° (H)	$\approx 29.30^\circ$ (W) × 19.21° (H)
The psychometric attributes scaled	Total image quality	Total image quality Brightness Saturation Naturalness Preference
Averaged ambient lighting	233 lux, ≈ 6500 K	233 lux, ≈ 6500 K

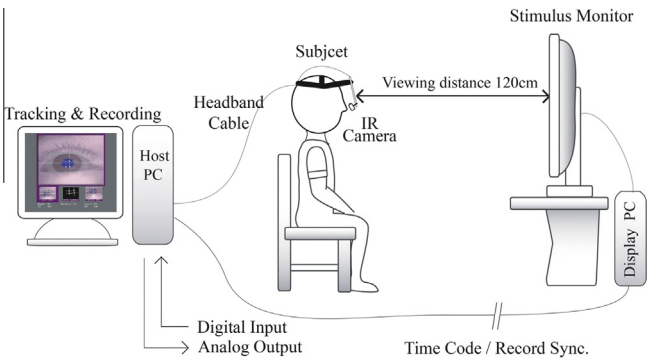


Fig. 1. The eye tracking apparatus set for experimental condition.

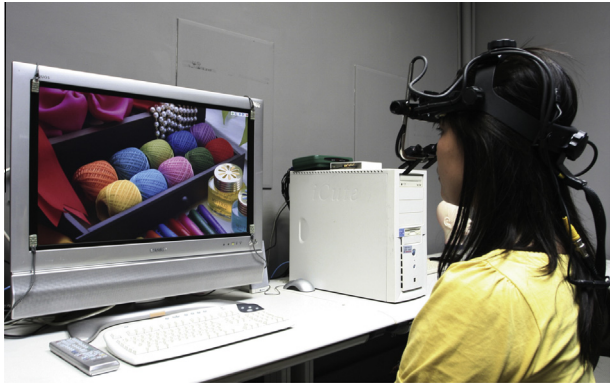


Fig. 2. The real condition of eye tracking experiment.

2.2.3. Experimental Images

The experimental images were selected from ISO standards (ISO 12640-1 (1997), ISO 12640-2(1997), ISO 12640-3(2004), ISO 12640-3(2007)), and some were collected from the Kodak Lossless True Color Image Suite [21]. All of these images were first categorized into three groups: portraits, landscapes with architectural images, and indoor multiple-object images. The experimental images were then selected from each category for the current experiment, as shown in Fig. 3. The size of an image was 1280 × 768 (pixels). The images were randomly displayed during each observation session against a background of a mid-gray color with an L^* of approximately 60.

2.3. Experiment II

Experiment II was performed to test the repeatability of the eye-tracking results. The environment for Experiment II was the same as that of Experiment I.

2.3.1. Display

A different display 40 in. in size (73.26° and 40.52° in the horizontal and vertical directions of full frame) was used. The viewing distance on the single image was adjusted to have the same visual field (29.30° and 19.21° in the horizontal and vertical directions, respectively).

2.3.2. Observers and experimental images

Fifteen observers (8 males and 7 females, whose average age was 27) participated in the experiment. Two of the subjects also took part in Experiment I. Fig. 4 shows the six images used, three of which were used in Experiment I.



Fig. 3. The images used in Experiment I.



Fig. 4. The images used in Experiment II.

Five assessment scales were used to assess image quality. These were selected from a pilot study in which 33 design students were asked to write down words to describe image quality. Altogether, ten attributes were accumulated. The five most frequently used attributes were chosen: 'total image quality', 'brightness', 'saturation', 'naturalness', and 'preference'. These attributes were used in Experiment II.

2.4. Experimental procedure

Fig. 5 shows the experimental procedure for both experiments, which was divided into 3 stages. In Stage I, after observers reported to the laboratory and passed the color vision test, they read the instructions. Each observer was then asked to look at 9 randomly

displayed fixation points on the display to calibrate the eye tracker. In Stage II, subjects viewed the experimental images arranged in a random sequence. Between two consecutive images, a calibration (drift correction) was performed to ensure the correct viewing position. In Stage III, subjects judged different scales for describing the image using a 7-point categorical judgment method for each image, where '1' was the lowest category, and '7' was the highest category.

2.5. Assessment scales

In Experiment I, each participant assessed eleven images using only one scale: 'total image quality.' The item for total image quality summarized the appearance of the image, including color,

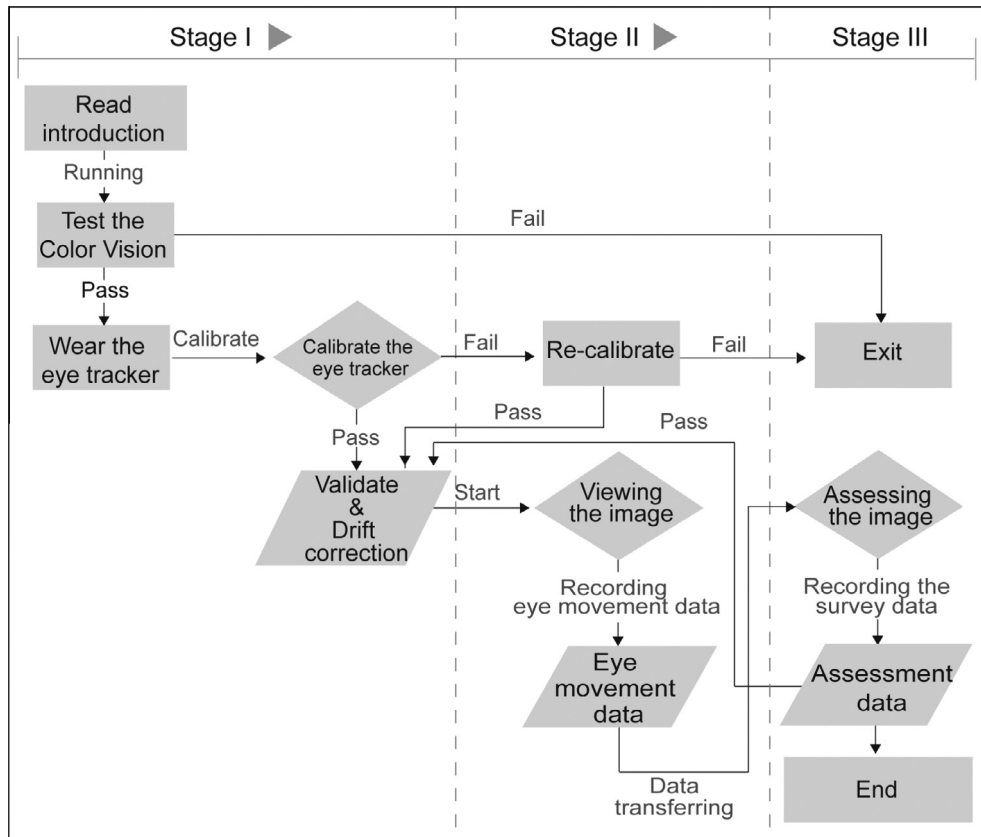


Fig. 5. The experimental procedure.

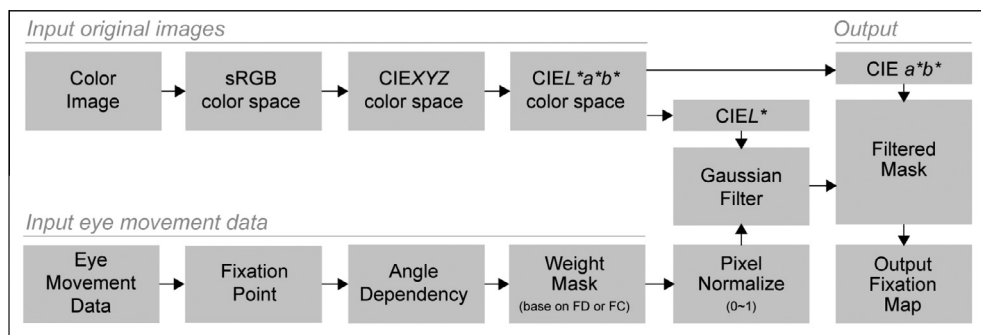


Fig. 6. The workflow to calculate Fixation Map.

lightness, tone, sharpness, and resolution. In Experiment II, six images were judged in terms of five scales ('total image quality', 'brightness', 'saturation', 'naturalness' and 'preference').

3. Data analysis

Henderson and Hollingworth [17] reported two important measures for viewing an image when studying eye movement. One is 'where to see', i.e., the fixation position tends to be centered during scene viewing, and the other is the 'how long to see', i.e., the fixation position tends to remain centered at a particular location in a scene [19]. In this study, the 'fixation duration (FD)' and 'fixation count (FC)' from the eye-tracking system were used, corresponding to 'how long' and 'how often', respectively. The following variables were considered: viewing angle dependency, observer variability, and number of observers. All of the results were reported in terms of CIELAB color differences [20]. Fig. 6 shows the workflow used to

obtain the final ROI description (filtered mask and fixation map) from the original image and eye-movement data.

In Fig. 6, each pixel of an image started as sRGB data (the earlier monitor was determined to be a sRGB monitor). It was then transformed to XYZ and then CIELAB L^* , a^* and b^* values. The L^* data were used to construct a filtered mask and fixation map with the visual data from the eye-tracking system.

The eye-movement data were first analyzed in a square for each fixation point, corresponding to a particular visual field size per unit of angle. Many fixation points were then collected from each observer's eye positions, which were used to produce a fixation map for each image.

3.1. Construction of the fixation map

The weight (w) of the fixation map was constructed by each pixel's fixation duration (FD) or fixation count (FC) relative to the

maximum value of the image considered. Finally, the L^* value of a pixel (i, j) was calculated using Eq. (1) by multiplying the L^* value of the original image by the weight of the pixel in question.

$$L_{ij}^* = L'_{ij} \times w_{ij} \quad (1)$$

where L'_{ij} is the L^* value for pixel (i, j) and w_{ij} is defined by Eq. (2) and

$$w_{ij} = \frac{I_{ij}}{I_{max_{ij}}} \quad (2)$$

where I_{ij} is the frequency of FD or FC for pixel (i, j) and I_{max} is the peak value of the frequency of the pixel in question.

3.2. Gaussian filter testing

A fixation map was filtered by a Gaussian filter given in Eq. (3) to construct a blurred mask of the fixation on each image.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

Here, x and y are the distances from the central point to the horizontal and vertical axes, respectively, and σ is the standard deviation of a Gaussian distribution.

3.3. Intensity of the Gaussian filter

The intensity of the filter was controlled using the standard deviation (σ) of a Gaussian filter (see Eq. (3)). A pilot study was conducted to investigate the effect of σ values. The fixation maps that were generated using different σ values were established from 1 to 96 at an interval of 5. Figs. 7 and 8 show the ΔL^* (or ΔE_{ab}^*) between the value of the two fixation maps (one has a σ value of 1, and the other has a different σ value for images in Experiments I and II, respectively). The results presented in both figures showed that the ΔL^* values were stabilized for a σ value equal to or greater than 30 regardless of which image was used. In other words, the filtered images would have only slight differences in their appearance for $\sigma > 30$. Thus, a σ of 30 was used in the following data analysis.

The variation between two images was analyzed using the ΔL^* formula (see Eq. (4)):

$$\Delta L_{mn}^* = \sqrt{(L_m - L_n)^2} \quad (4)$$

where L_m is the $CIE L^*$ value of each pixel on the one image and L_n is the $CIE L^*$ value of each pixel on another image.

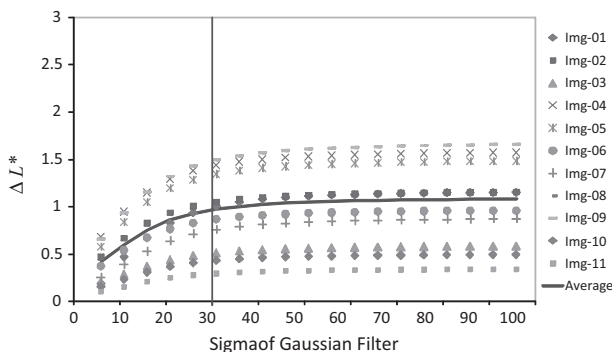


Fig. 7. Results in ΔL^* unit calculated between different intensity in σ unit of Gaussian filter and σ of 1 for all Experiment I images.

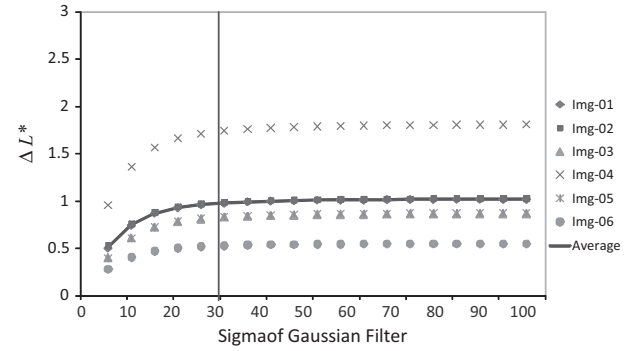


Fig. 8. Results in ΔL^* unit calculated between different intensity in σ unit of Gaussian filter and σ of 1 for all Experiment II images.

4. Results and discussion

4.1. Methods for counting fixations and defining visual sizes

The parameters affecting the distribution of spatial frequency on a fixation map were the intensity of the fixation point and the size of the viewing field. As noted above, two types of data were used to define fixations for each ROI. The FD parameter summarizes the fixation distributions based on fixation time for each pixel. The FC parameter summarizes the fixation distributions based on the fixation area count for each pixel. For each image, ΔL^* was calculated to compare the fixation maps using FD and FC data for each pixel. The results were averaged to express the perceived differences between two types of data.

The size of the viewing field when observers view the ROI in an image could also affect the fixation map constructed. Theoretically, the fovea vision corresponds to approximately 2° of the visual field in the retina, which is responsible for color perception when viewing images, reading, and other activities for which visual detail is of primary importance. The region around the point of fixation is projected onto the fovea, where the retinal image is perceived with the highest resolution. In this study, the fixation maps were constructed with sampling field sizes of 1° , 2° and 4° , respectively. Tables 2 and 3 give the ΔL^* values, including FD and FC data, of three visual fields for each image in Experiments I and II, respectively.

The results presented in Table 2 show that FD and FC data are in good agreement; they possess small ΔL^* values of approximately 0.7 for all three visual fields. Additionally, there are only slight differences among the three different viewing field sizes. This finding implies that FD and FC will yield similar results regardless of which visual size is used. Comparing the results from different experiments, Experiment II images had an approximately 130% larger FD and FC difference than Experiment I, possibly due to the slightly larger display size used in Experiment II. The results presented in Tables 2 and 3 are also plotted in Figs. 9 and 10 for Experiment I and II images, respectively.

The majority of the 1° visual field ΔL^* values are the highest (larger difference between FD and FC data), and the 4° visual field ΔL^* values are the lowest. However, the differences are quite small. All of the images, including facial content, had a smaller value than the other images, indicating that the ROIs found were more accurate when the image involved human subjects.

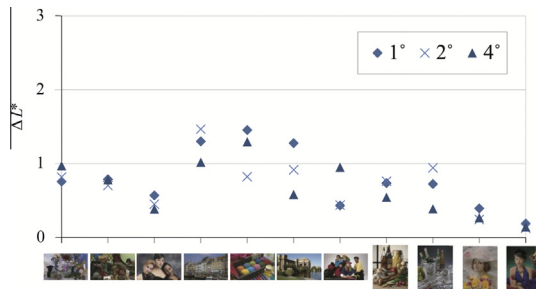
Pairs of results (1° versus 2° , 1° versus 4° , and 2° versus 4°) were used to investigate the differences caused by field of view. Tables 4 and 5 list the ΔL^* values for each pair, with different sizes for FD and FC data for Experiment I and II images, respectively. The 1° and 4° images have a greater ΔL^* value than observed for the other pairs. In both experiments, the ΔL^* values calculated for 2° and 4°

Table 2Results in ΔL^* values between FD and FC data on each visual field for all of the Experiment I images.

Viewing fields	1	2	3	4	5	6	7	8	9	10	11	Ave.
1°	0.755	0.786	0.567	1.299	1.453	1.275	0.430	0.734	0.722	0.391	0.185	0.782
2°	0.818	0.703	0.449	1.465	0.823	0.917	0.437	0.761	0.942	0.244	0.125	0.698
4°	0.972	0.780	0.381	1.018	1.294	0.581	0.950	0.543	0.385	0.265	0.141	0.664

Table 3Results in ΔL^* values between FD and FC data on each visual field for all of the Experiment II images.

Viewing fields	1	2	3	4	5	6	Ave.
1°	0.808	2.030	1.071	1.919	0.935	1.519	1.380
2°	0.713	1.179	0.543	1.483	0.542	1.020	0.913
4°	0.785	0.573	0.487	0.964	0.558	0.547	0.652

**Fig. 9.** The ΔL^* values calculated between FD and FC data for each visual field of all Experiment I images.**Fig. 10.** The ΔL^* values calculated between FD and FC data for each visual field of all Experiment II images.

are larger than those for 1° and 2°. According to Figs. 9 and 10, 2° data are in agreement with the results from the other visual fields. The 2° visual angle refers to a small area of the retina that has a high visual acuity in a human eye. Moreover, fixation duration (FD) could directly affect the time that it takes to look at a small

Table 4Results in ΔL^* units between each pair of visual fields based on FD and FC data for the Experiment I images.

Pairs of viewing fields	Number of images											Ave.
	1	2	3	4	5	6	7	8	9	10	11	
<i>Fixation duration method</i>												
1° versus 2°	1.837	2.370	1.103	3.281	2.837	1.706	1.542	4.121	1.760	0.762	0.705	2.002
1° versus 4°	4.772	4.787	3.123	5.865	4.571	3.583	5.388	6.294	3.600	2.275	1.828	4.190
2° versus 4°	3.237	2.684	2.274	3.023	2.129	2.207	4.091	3.333	2.388	1.759	1.330	2.587
<i>Fixation count method</i>												
1° versus 2°	3.062	3.550	2.109	4.776	2.748	2.866	2.746	6.583	3.469	1.596	1.959	2.092
1° versus 4°	6.619	7.019	4.972	7.734	6.497	5.071	7.913	10.773	6.940	5.183	4.605	4.553
2° versus 4°	4.363	4.348	3.831	3.992	4.833	3.125	6.360	5.489	4.101	4.196	3.576	2.911

Table 5Results in ΔL^* units between each pair of visual fields based on FD and FC data for the Experiment II images.

Pairs of viewingfields	Number of images						Ave.
	1	2	3	4	5	6	
<i>Fixation duration method</i>							
1° versus 2°	1.187	1.555	1.206	2.881	1.126	0.916	1.479
1° versus 4°	2.849	2.787	2.604	4.619	3.341	1.555	2.959
2° versus 4°	2.147	1.926	1.941	2.846	2.684	0.994	2.090
<i>Fixation count method</i>							
1° versus 2°	1.402	1.776	1.202	3.208	1.209	1.204	1.667
1° versus 4°	3.278	2.711	2.282	4.718	3.338	1.645	2.995
2° versus 4°	2.430	1.800	1.639	2.780	2.655	0.937	2.040

Table 6

Results of observer variation between Experiments I and II.

Analysis	Experiment I	Experiment II
Between-observer variability (BOV)	5.141	5.192
Interobserver variability (IOV)	3.630	2.480

area in the image stimuli. Thus, the 2° visual field of FD data was used for the following data analysis.

4.2. Observer variation

Observer variability was analyzed in two ways: between-observer variability (BOV) and interobserver variability (IOV).

4.2.1. BOV and IOV results

As noted above, each observer's data were presented in the form of a fixation map for the 2° field FD data. These images were used to calculate the BOV and IOV using the ΔL^* unit. The ΔL^* value for the IOV was calculated to compare the fixation map of the mean observer data to that of each individual observer. Finally, the ΔL^* value calculated from all observers was averaged with the IOV. The BOV was calculated for every combination of the fixation maps for two observers. Table 6 summarizes the IOV and BOV results.

Table 6 shows that the BOV has larger variability than the IOV. The IOV results are consistently smaller than the BOV results by a

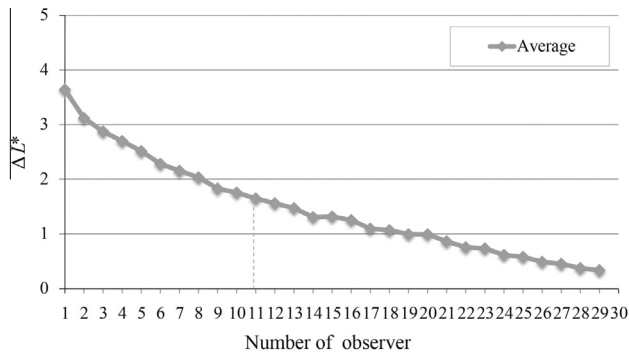


Fig. 11. Results in ΔL^* units calculated between mean data from different number of observers and from 29 observers (Experiment I).

Table 7

Results of differential scale testing on Experiment II (ΔL^*).

Scale	Brightness	Saturation	Naturalness	Preference
Average	1.503	1.499	1.431	1.509
Minimum	0.785	0.591	0.515	0.577
Maximum	3.191	2.730	2.441	2.443
SD	0.842	0.865	0.830	0.647

factor of approximately 1.5. These values represent the typical observer variability involved in the eye-tracking experiment.

4.2.2. Number of observers

The next data analysis was conducted to determine the number of observers required to obtain reliable data, taking the mean results from all 30 observers as the standard. This approach was

tested by randomly selecting subsets of observers; each subset contained between 1 and 29 observers. For example, a subset of 15 observers was randomly selected, and their fixation map results were averaged. The mean from the subset and the full set (30 observers) were compared in terms of their ΔL^* values. The process was repeated 10 times, and the results were averaged. Fig. 11 summarizes the resulting ΔL^* values for different subsets. As expected, an increase in the number of observers in a subset agrees better with the full set of results.

The results of the one-way analysis of variance (ANOVA) showed a significant difference depending on the number of observers used ($F_{(28, 319)} = 12.269, p < 0.001$) [22]. A post-hoc Tukey HSD test [22] was also conducted. The results from 11 observers, which corresponded to a ΔL^* value of 1.64, were not significantly different than the results from the full set of 30 observers.

4.3. Differential quality scales

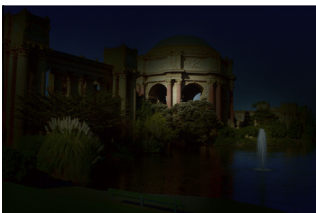
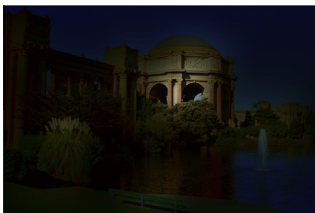


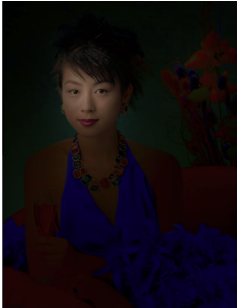
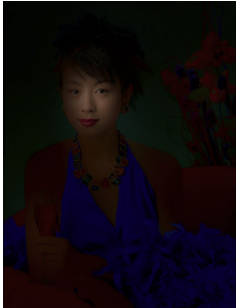
In Experiment II, five scales were used to evaluate the quality of each image. The experiment was conducted one scale at a time so that five different fixation maps (one scale each) were obtained. To test which scale is closest to the total image quality, the fixation maps for each scale were compared to the 'total image quality' scale, again expressed in terms of the ΔL^* unit. The results are summarized in Table 7. The ΔL^* values for each scale are quite similar, i. e., all scales had a mean ΔL^* value of approximately 1.5. In other words, all 4 scales contribute equally to the image quality.

4.4. Repeatability between results from Experiments I and II

The next test was performed to determine the repeatability of the data. The three images (landscape, multicultural faces, and the Japanese lady) used in both experiments were investigated.

Table 8

Results of a fixation map compared between Experiments I and II (ΔL^*).

Experiment	I	II	ΔL^*
Landscape			1.810
Portrait			1.366
Lady			0.933

Again, the ΔL^* value was used to indicate repeatability and was calculated for the two mean fixation maps in Experiments I and II to assess the total image-quality tasks. The results (in ΔL^* units) and the fixation maps are shown in Table 8. The repeatability performances were 1.8, 1.4 and 0.9 for the landscape, portrait, and Japanese-lady images, respectively. These differences are smaller than the IOV (approximately 2.5 ΔL^* units) found earlier, indicating that the repeatability between the two sets of experimental results is acceptable.

The magnitude of the color differences in Table 8 can be explained by the fact that when viewing landscape images, observers viewed the entire image on a global basis, ranking objects as equally important so that no particular object stood out. Additionally, familiar objects (or uniform scene areas) such as blue skies, grass and foliage did not attract observers' attention and were unimportant, as reported by other studies [16,23]. This could be partially due to the large uniform scene area that these familiar objects occupied [24]. In a portrait image including one or more people, the observers focused on human faces, especially the eyes. In addition, observers appear to spend more time looking at Asian

faces compared to those of other races, possibly because only Asian observers were involved in this experiment. These trends are highly consistent for both sets of experiments.

4.5. Fixation map or ROIs

Detailed inspection was carried out by comparing the fixation maps. Images can be divided into three types, which are illustrated in Fig. 12: portrait/animal images, object images, and landscape/building images. For the first type of images, the observers focused on the portrait's faces and the parrot's eyes. Some of the portrait images included different numbers of human subjects. In all cases, the facial areas attract more attention. Fig. 12 also shows the results for artificial objects shown against a background. The large, more uniform background was frequently ignored. The results of the fixation map also showed that the central region of the image seemed to be the most alluring area. The observers tend to view the landscape/building images globally, with no particular region of interest. However, blue skies, grass and foliage did not draw observers' attention, possibly because these objects cover

(A) Portrait & Animals Images



(B) Objects Images



(C) Landscape & Building Images



Fig. 12. Fixation maps of different kind of images.

Table 9

Summary of the experimental results.

ΔL^*	Experiment I	Experiment II
Visual field: 1° versus 2°	2.0	2.1
FD versus FC data at 2°	0.7	0.9
Interobserver variability (IOV)	3.6	2.5
Repeatability between experiments I and II	–	1.4

relatively homogenous regions, as reported by other studies of landscape images [25]. Observers also tended to focus on the image center when the image is more complex, containing items such as buildings, boards, and cross lines (see Fig. 12).

5. Conclusions

In this study, two experiments were performed to examine the role of ROIs in the task of assessing image quality for 17 images using an eye-movement system. The results can be described in terms of a fixation map, and the distribution of eye movements between images can be directly analyzed in terms of CIELAB color differences. The results from two experiments were highly consistent. This implies that the eye-tracking technique can provide robust results. Additionally, it is easy to understand the experimental uncertainty in terms of color differences.

Table 9 summarizes the experimental uncertainties in terms of ΔL^* values, including visual field, FD and FC method, interobserver variability and repeatability. The results of FD and FC are quite similar. The results showed that the typical IOV result, representing the typical observer variability, is approximately 3 ΔL^* units. The repeatability for images used in both experiments is good and was smaller than IOV. This result concluded that eye-tracking is robust for studies that evaluate image quality.

In different types of images, there was a clear trend that the observers tend to focus on either human faces or animal eyes. Vu et al. examines visual fixation patterns in subjects performing an image quality assessment task, with similar results [26]. This study also shows that observers tend to ignore the blue sky, grass and foliage in the landscape images; they nearly always pay attention to the image's main objects, such as a building or a fountain. The results were similar to those found by Santella and DeCarlo [16]. Observers could not clearly identify ROIs when the images included too many objects; thus, they almost always focused on the central region of the images. The central region attracted more attention than the background and the other objects in the images studied.

As noted earlier, Privitera et al. [15] and Henderson and Hollingworth [17] paved the way for a new wave of ROI research. This study continued those earlier studies by focusing on a more detailed examination of ROIs. Eye-movement measurements provide a valuable method for describing imaging science in terms of ROIs. The fixation map method described here is effective in correlating ROIs to an image. The color-difference or lightness-difference value calculated between two fixation maps is both easy to understand and highly effective because it computes the difference only based on ROIs, which receive the most attention.

Acknowledgments

The authors would like to express thanks for the NSC scholarship from Taiwan (ROC) that enabled the main author of this paper to study at the Colour, Imaging and Design Center, at the University of Leeds. The authors would also like to thank Prof. Ronnier Luo

(School of Design, University of Leeds) for the suggestion at analysis method, and Prof. Stephen Westland (School of Design, University of Leeds) for comments that greatly improved the manuscript. Special thanks to anonymous reviewers for their so-called insights. The authors would like also immensely grateful to Ph.D. Pei-Li Sun (National Taiwan University of Science and Technology) and Ph.D. Wang-Chin Tsai (Fo Guang University) for their comments on an earlier version of the manuscript.

References

- [1] V. Loia, S. Sessa, Fuzzy relation equations for coding/decoding processes of images and videos, *Inform. Sci.-Inform. Comput. Sci.: Int. J.* 171 (2005) 145–172.
- [2] H. Nobuhara, W. Pedrycz, S. Sessac, K. Hirota, A motion compression/reconstruction method based on max T-norm composite fuzzy relational equations, *Inform. Sci.* 176 (2006) 2526–2552.
- [3] A. Nguyen, V. Chandran, S. Sridharan, Gaze tracking for region of interest coding in JPEG 2000, *Signal Process.: Image Commun.* 21 (2006) 356–377.
- [4] M.R. Civanlar, Content adaptive video coding and transport, in: *Proc. Signal Processing and Communications Applications Conference of the IEEE 12th*, 2004.
- [5] K. Oda, A. Yuuki, T. Teragaki, Evaluation of moving picture quality using the pursuit camera system, *Euro Display* 6 (2002) 115–118.
- [6] S.Y. Choi, M.R. Luo, M.R. Pointer, P.A. Rhodes, Investigation of large display colour image appearance. I: Important factors affecting perceived quality, *J. Imaging Sci. Technol.* 52 (2008). 040904-1–11.
- [7] S.Y. Choi, M.R. Luo, M.R. Pointer, P.A. Rhodes, Investigation of large display colour image appearance. II: The influence of surround conditions, *J. Imaging Sci. Technol.* 52 (2008). 040905-1–9.
- [8] S.Y. Choi, M.R. Luo, M.R. Pointer, P.A. Rhodes, Investigation of large display colour image appearance III: Modelling image naturalness, *J. Imaging Sci. Technol.* 53 (2009) 031104-1–031104-12.
- [9] H.R. Sheikh, A.C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2006) 430–444.
- [10] G. Ginesu, F. Massidda, D.D. Giusto, A multi-factors approach for image quality assessment based on a human visual system model, *Signal Process.: Image Commun.* 21 (2006) 316–333.
- [11] A.J. Calabria, M.D. Fairchild, Perceived Image Contrast and Observer Preference I: The Effects of Lightness, Chroma, and Sharpness, *Manipulations on Contrast Perception*, Munsell Color Science Laboratory, Rochester Institute of Technology, 2003.
- [12] J.E. Sheedy, R. Smith, J. Hayes, Visual effects of the luminance surrounding a computer display, *Ergonomics* 48 (2005) 1114–1128.
- [13] E.A. Fedorovskaya, H. Ridder, F.J. Blommaert, Chroma variants and perceived quality of color images of natural scenes, *Color Res. Appl.* 22 (1996) 96–110.
- [14] C.M. Privitera, L.W. Stark, Algorithms for defining visual regions-of-interest: comparison with eye fixations, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000) 9970–9982.
- [15] C.M. Privitera, T. Fujita, D. Chernyak, L.W. Stark, On the discriminability of hROIs, human visually selected regions of interest, *Biol. Cybern.* 93 (2005) 141–152.
- [16] A. Santella, D. DeCarlo, Robust clustering of eye movement recordings for quantification of visual interest, in: *Proc. Eye Tracking Research and Applications Symposium (ETRA)*, 2004.
- [17] J.M. Henderson, A. Hollingworth, Eye movements during scene viewing: an overview, in: G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception*, Elsevier, Oxford, 1998, pp. 269–294.
- [18] D.S. Wooding, Fixation maps: quantifying eye-movement traces, in: *Proc. Eye Tracking Research & Applications Symposium (ETRA)*, 2002.
- [19] A.T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag, London, 2003.
- [20] CIE, *Colorimetry*, 15, CIE, Vienna, 2004.
- [21] Kodak, Kodak Lossless True Color Image Suite, 2007. <<http://r0k.us/graphics/kodak/>> (retrieved from August 2007).
- [22] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tatham, *Multivariate Data Analysis*, seventh ed., Prentice Hall, NJ: Englewood Cliffs, 2009.
- [23] J.M. Henderson, J.R. Brockmole, M.S. Castelano, M. Mack, Visual saliency does not account for eye movements during visual search in real-world scenes, in: R.P.G. van Gompel, M.H. Fischer, W.S. Murray, R.L. Hill (Eds.), *Eye Movements: A Window on Mind and Brain*, Elsevier Ltd., Oxford, UK, 2007, pp. 537–561.
- [24] J.M. Henderson, A. Hollingworth, High-level scene perception, *Annu. Rev. Psychol.* 50 (1999) 243–271.
- [25] J.M. Henderson, Human gaze control during real-world scene perception, *TRENDS Cogn. Sci.* 7 (2003) 498–504.
- [26] C.T. Vu, E.C. Larson, D.M. Chandler, Visual fixation patterns when judging image quality: effects of distortion type, amount, and subject experience, in: *Proc. Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, 2008, pp. 73–76.