Using gaze information to improve image difference metrics

Marius Pedersen, Jon Yngve Hardeberg and Peter Nussbaum; Gjøvik University College; Gjøvik, Norway

ABSTRACT

We have used image difference metrics to measure the quality of a set of images to know how well they predict perceived image difference. We carried out a psychophysical experiment with 25 observers along with a recording of the observers gaze position. The image difference metrics used were CIELAB ΔE_{ab}, S-CIELAB, the hue angle algorithm, iCAM and SSIM. A frequency map from the eye tracker data was applied as a weighting to the image difference metrics. The results indicate an improvement in correlation between the predicted image difference and the perceived image difference.

Keywords: Image difference metrics, Eye tracking, CIELAB ΔE^*_{ab}, S-CIELAB, SSIM, Hue angle, iCAM.

1. INTRODUCTION

Image difference metrics have been proven shortcoming when it comes to predict perceived image difference. Information on where we look in images have been proven useful many different fields, also in image quality. Areas where observers spot a difference in images should be weighted higher in a pixel-by-pixel difference. We use information based on where observers gaze at images as a weight for different image difference metrics. If an improvement is found, automatic region-of-interest algorithms simulating observers gaze could be used to improve image difference metrics without doing an eye tracker experiment.

2. STATE OF THE ART

From the CIELAB ΔE_{ab} in 1976 many different image difference metrics have been published, in a number of studies it has been shown that the results does not always correlate with the perceived image quality.[1,2]

2.1 Hong and Luo's hue angle metric

This algorithm by Hong and Luo[3] is based on the known fact that systematic errors over the entire image is quite noticeable and unacceptable. The proposed algorithm is based on some conjectures:

- Pixels or areas of high significance can be identified and a suitable weight allocation can be found.
- Larger areas of the same color should be weighted higher.
- Larger color difference between the pixels should get higher weights.
- Hue is an important color percept for discriminating colors within the context.

The proposed algorithm creates a histogram based on the hue angle, and sorts this ascending so weights can be applied to different sections of the histogram. The overall color difference is then calculated by multiplying the weighted hue angle for every pixel with the color difference pixel-by-pixel.

2.2 SSIM

The SSIM proposed by Wang et al.[4] attempt to quantify the visibility between a distorted image and a reference image. The algorithm defines the structural information in an image as those attributes that represent the structure of the objects in the scene, independent of the average luminance and contrast. The index is based on a combination of luminance comparison, contrast comparison and structure comparison. The comparison is done for local windows in the image, and the overall image quality is the mean of all these windows.

Human Vision and Electronic Imaging XIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6806, 680611, © 2008 SPIE-IS&T · 0277-786X/08/$18

2.3 S-CIELAB

The S-CIELAB model was designed as a spatial pre-processor to the standard CIE color difference equations,[5] to account for complex color stimuli such as halftone patterns.[6,7] The S-CIELAB has two goals, first it would like to apply a spatial filtering operation to the color image data, to simulate spatial blurring by the human visual system. Second, in large uniform areas, they would like the extension to be consistent with the basic CIELAB calculation. The image data are transformed into opponent-color space, each of these opponent-colors are convolved with a kernel. The filtered representation is transformed to CIE-XYZ representation, this resulting representation includes both spatial filtering and the CIELAB processing. The difference is summarized using ΔE_s like the conventional CIELAB.

2.4 iCAM

The iCAM model was proposed by Fairchild and Johnson.[8] iCAM was built upon previous research in many fields among uniform color space[9] because of the hue-linearity,[10] image surround importance,[11] image difference and image quality measurement algorithms.[12,13] This model takes the tristimulus values and transform them into RGB values using Von Kries adaptation identical to the one found in CIECAM02.[14] Further the adopted signals are transformed into the IPT color space.[9] The adapting and the surround luminance level is then used to allow for the prediction of various appearance phenomena.

2.5 Region-of-interest and image difference metrics

Bando et al.[1] used regions based on criteras from the observers were used to compute areas used to calculate image difference. The conclusion by the authors was no correlation between the average pixel-by-pixel difference and perceived image quality. Morovic and Sun[15] did research on differences in color image reproduction experiments, they discovered that more than 50% of the important errors were perceived in parts of the image rather than in the entire image. They also found out that size and location of an object within an image do not have a strong impact on whether it's difference will be jugded to important. In the following research Morovic and Sun[16] found out that 80% of differences were reported due to changes in lightness, colorfulness and hue.

Bai et al.[17] evaluated S-CIELAB on images produced by Retinex by using gaze information. The average S-CIELAB color difference was weighted by the frequency map from the gazing information on the whole image and over the gazing areas. The results from this research shows that frequency distribution of gazing area in the image gives important information on the evaluation of image quality.

3. EXPERIMENT SETUP

3.1 Images

4 different images have been used in this experiment (Figure 1). All the images have been reproduced with

(a) Girl reproduced with permission from Grafisk Assitanse AB.

(b) Tore reproduced with permission from Se og Hør.

(c) JP from ISO.

(d) Cartoon reproduced with permission from Trond Viggo Bjerke.

Figure 1. Images used in the experiment.

different changes in lightness. Each scene has been altered in 4 directions globally and 2 different regions 2 directions locally, resulting in a total of 36 images. Changes was made in CIELAB color space in only the L^* channel in 16 bit, the global changes were plus and minus 3 and 5 ΔE_{ab} in lightness. The profile from the display was also applied to the images.

3.1.1 Regions

In the images with only changes to regions in the image, the value used to increase and decrease lightness in the whole image was applied only to the regions. I.e. the regions have been altered to the same degree as the same region for the globally changed images. The region in the images with only local changes is similar to the same region in the images with global changes. The white areas in Figure 2 were altered and the black were not.

(a) Girl back (b) Girl front (c) Tore back (d) Tore front (e) JP stripe (f) JP stripe2 (g) Cartoon sign (h) Cartoon t-shirt

Figure 2. Regions, only white pixels have been altered.

3.2 Psychophysical experiment setup

The psychophysical experiment was done on a calibrated CRT monitor, LaCIE electron 22 blue II, in a gray room. The observer was seated approximately 80 cm from the screen,[18] this distance gives the eye tracker a head tracking area of about 40x40 cm. The eye tracker used was a SMI iViewX RED,[19] a contact free gaze measurement device. The light was dimmed and measured approximately 17 lux in front of the screen, this is within the CIE recommendation of 64 lux.[20] Observers were shown 3 images at once during the experiment, one original image in the middle and two different reproductions on each side. Observers were told to choose the image most similar to the original.

The z-scores are based on Thurstone's law of comparative judgement with 95% confidence intervals.[21] The error bars are computed as

$$\bar{X} \pm \frac{\sigma}{\sqrt{N}}$$

where $\bar{X}$ is the z score, σ is the the standard deviation and N is the size of the sample. For this experiment this is the number of observers multiplied with 2, because each image was shown twice for consistency.

4. RESULTS

4.1 Results from psychophysical experiments

The results from the 25 observers can be seen on Figure 3, this image shows the mean results over the 4 scenes for the images with global lightness changes of minus 3 ΔE_{ab} (LM3), minus 5 ΔE_{ab} (LM5), plus 3 ΔE_{ab} (LP3) and plus 5 ΔE_{ab}(LP5). From the results we can see that LM3 and LP3 have the same visual distance from the orignal, the same with LM5 and LP5. The observers can see a big difference between lightness 3 ΔE_{ab} and 5 ΔE_{ab} even though the difference is only 2 ΔE_{ab}.

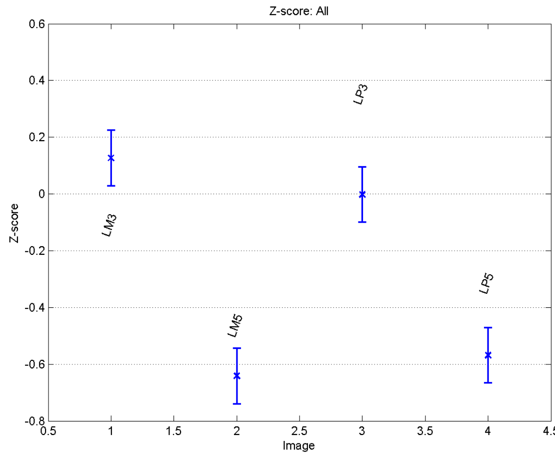


Figure 3. Z-score for global lightness changes (LM3/5 and LP3/5)

4.1.1 Girl

For the girl image, we can see that BACK LM3 (background lightness minus 3 ΔE_{ab}), BACK LP3 (background lightness plus 3 ΔE_{ab}) and FRONT LM3 (front lightness minus 3 ΔE_{ab}) score the highest , and have the same visual difference from the original. LM3, LP3 and FRONT LP3 score approximately the same, while LM5 and LP5 score significantly lower than the rest. From the results shown in Figure 4 smaller changes in background have a smaller visual difference from the original than global changes. In the globally changed images both the semantic regions and the almost uniform background is altered, both of these were stated as important by the observers and therefore these also have the largest visual difference from the original.

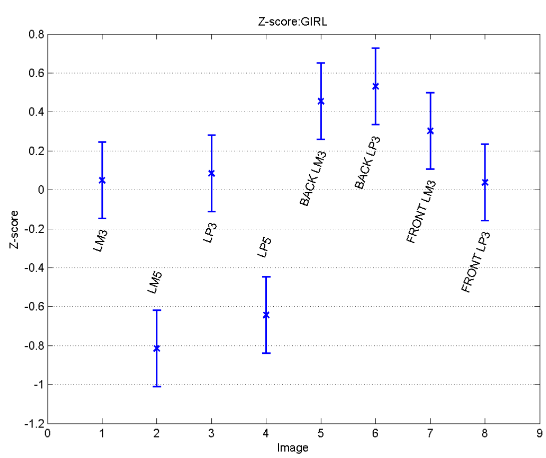


Figure 4. Z-score for GIRL

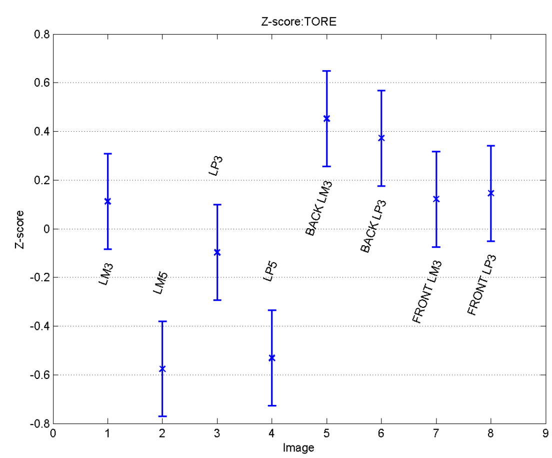


Figure 5. Z-score for TORE

4.1.2 Tore

For the Tore scene BACK LM3 scores the highest (Figure 5), but it has the same distance from the orignal as BACK LP3, FRONT LM3/LP3 and LM3. Observers found images with brighter or darker background to be closer to the original than brighter global images. The background in this scene is non-uniform making it difficult to spot a difference here, observers will therefore look at areas in the foreground to find a difference. Both the LM5 and LP5 score significantly lower than the rest, these also being the images with the largest modifications.

4.1.3 JP

In this scene observers rated low global changes and the vertical stripe on the right side of the image to be closest to the original (Figure 6), while higher global changes together with the horizontal stripe in the middle of the image have a larger visual difference from the original. The horizontal stripe has a large visual difference from the original according to the observers, but a small ΔE^*_{ab} change. The small changes made in this image have a high visibility in central parts of the image, resulting in artifacts as new edges. Therefore observers rate other images closer to the original.

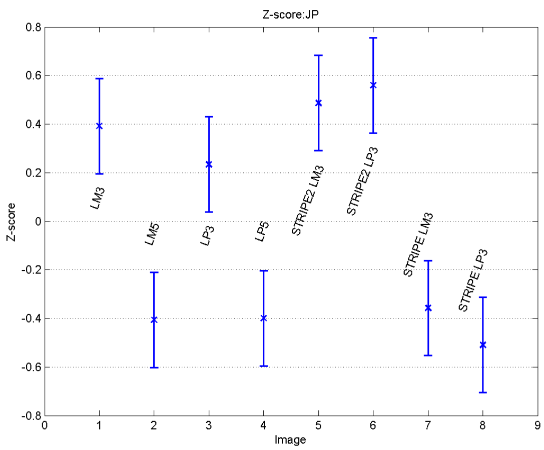


Figure 6. Z-score for JP

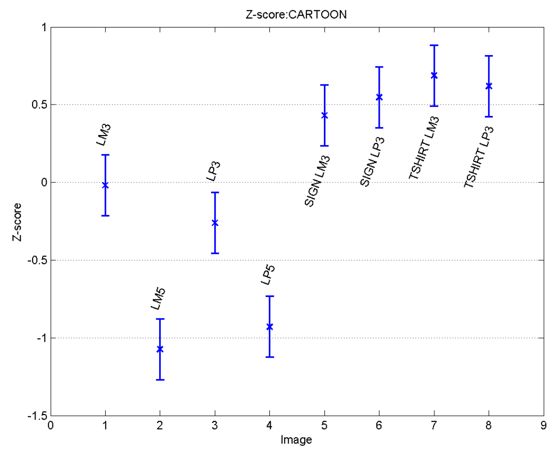


Figure 7. Z-score for CARTOON

4.1.4 CARTOON

Here Tshirt LM3 has the lowest visual difference from the original (Figure 7), but the difference is not distinguishable from Tshirt LP3, SIGN LM3 or SIGN LP3. LM3 and LP3 have a larger visual difference from the orignal than the previous. LM5 and LP5 have a significantly larger visual difference than these again. In the globally changed images observers stated that they could easily see a difference in the uniform background, and therefore these images were rated with a larger difference to the orignal than the images with only small local changes.

4.2 Image difference metrics results

The image difference metrics used are ΔE_{ab}, S-CIELAB, iCAM, SSIM and the hue angle algorithm. Pearson's correlation coefficients for each scene is shown in Table 1.

Scene/Algorithm	S-CIELAB	iCAM	ΔE_{ab}	SSIM	Hue angle
1 (Girl)	0.94	0.63	0.93	0.61	0.96
2 (Tore)	0.96	0.57	0.96	0.50	0.97
3 (JP)	0.09	0.01	0.08	0.05	0.16
4 (Cartoon)	0.97	0.74	0.95	0.85	0.97

Table 1. Correlation coefficient r^2 for the different algorithms

4.2.1 ΔE_{ab}

The scores from the ΔE_{ab} can be seen in Figure 8. ΔE_{ab} scores a correlation coefficient (r^2) of 0.63, this indicates a correlation between the z-scores and the algorithm (Figure 8).

Images with only small regions altered scores low in the ΔE_{ab} and have high z-scores except for the 2 horizontal stripes in the JP scene. The low ΔE_{ab} values indicates that this image should score high in a psychophysical experiment, but this is the not the case. We can also see that the images altered with 5 ΔE_{ab} have very different z-scores. One of the Cartoon images (LM5) scores lower than -1 on the z-score and one of the JP images score approximately -0.4, this implies that the observers can differentiate between these images while the ΔE_{ab} cannot. The same for the images altered with 3 ΔE_{ab}, where we have a z-score difference of 0.6.

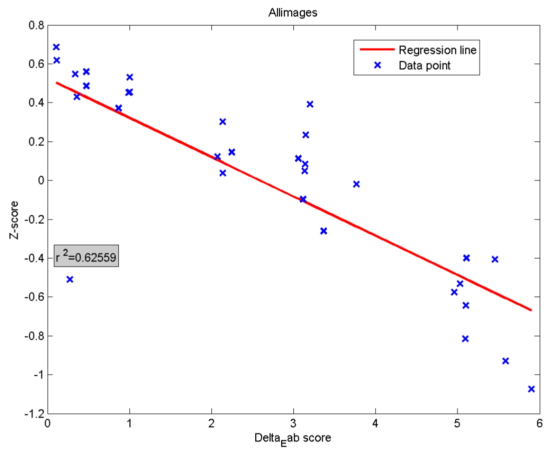


Figure 8. Z-score vs. ΔE_{ab}

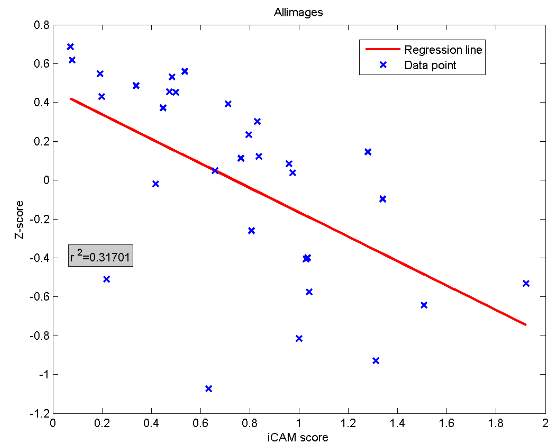


Figure 9. Z-score vs. iCAM

4.2.2 iCAM

iCAM shows a more scattered result as seen on Figure 9 than ΔE_{ab} in the previous section. Some of the images with a low iCAM score have a high z-score, but other images with a low iCAM score have a low z-score. We can see the same for iCAM regarding the two JP images with the horizontal stripe as in ΔE_{ab}. iCAM also has problems with the LM5 images, in 3 scenes this image is rated much better than LP5 even though observers rate them with the same visual difference from the original.

4.2.3 Hue angle

The hue angle algorithm gives a low score to images where only small regions have been altered, and a higher score to images with a global score (Figure 10). The two images with a horizontal stripe score close to 0 in this algorithm, but the z-score implies that these images should score higher. The hue angle algorithm differentiate well between the images with 3 and 5 ΔE_{ab}.

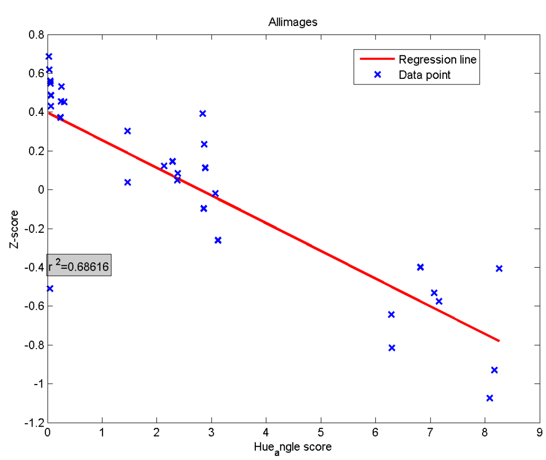


Figure 10. Z-score vs. hue angle

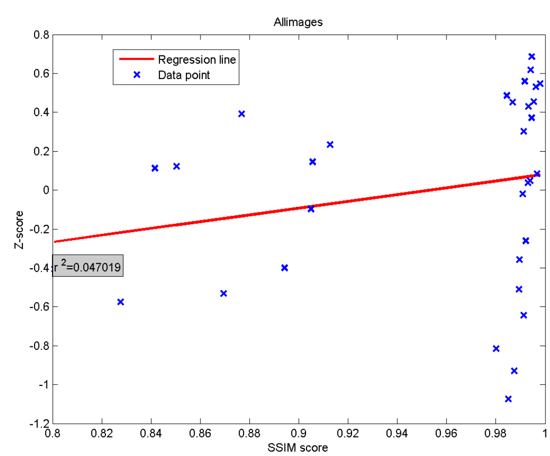
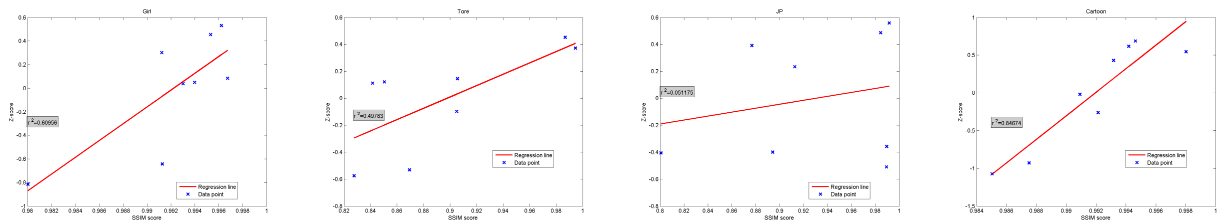


Figure 11. Z-score vs. SSIM

4.2.4 SSIM

The SSIM algorithm is made for grayscale images, the images used in this experiment have only been altered in lightness therefore the SSIM should give a valid result. From Figure 11 we can see that the Tore scene gives very different results, all from 0.82 to 0.99. The r^2 gives a score of only 0.05, the reason for this overall low score is the difference in the results between the scenes. We get a better view of the performance of SSIM by looking at the single scenes in Figure 12. We can see that the only scene SSIM performs well in is the Cartoon. In the JP scene 4 images with more or less the same z-score have been rated all over the scale.



(a) Girl (b) Tore (c) JP (d) Cartoon

Figure 12. SSIM for each scene

4.2.5 S-CIELAB

As we can see from Figure 13 the results are fairly linear, resulting in a r^2 of 0.61. There are some outliers, the most obvious are the horizontal JP stripe. These score low both on the z-score and S-CIELAB score. The images changed plus or minus 5 ΔE_{ab} are rated on the higher end of the scale in S-CIELAB, and the images with smaller regions are in the other end of the scale.

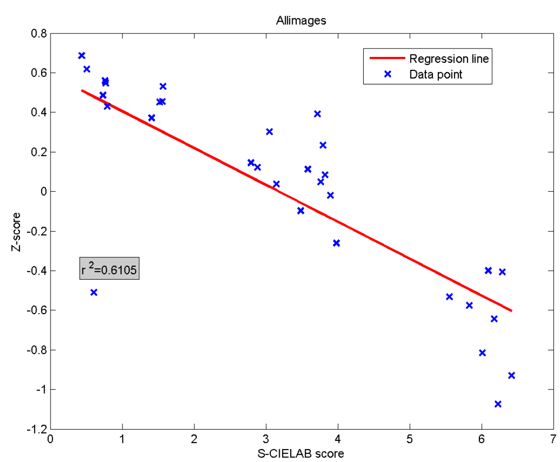


Figure 13. Z-score vs. S-CIELAB

4.2.6 Difference between the algorithms

To calculate the difference in the performance of the algorithms, the linear regression line for each scene for the different algorithms were computed. From this regression line the mean squared difference with a 95% confidence interval were calculated as a measure of performance for the algorithms together with the correlation coefficient r^2.

Figure 14 shows that S-CIELAB, ΔE_{ab} and hue angle algorithm scores better than SSIM. The hue angle algorithm has the best score, with only a little difference to the ΔE_{ab} and S-CIELAB. This is not surprising because of the familiarity between these algorithms, since both the S-CIELAB and hue angle are build upon the

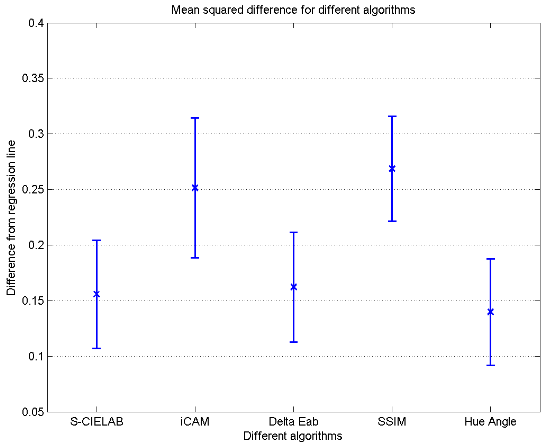


Figure 14. Mean squared difference for algorithms

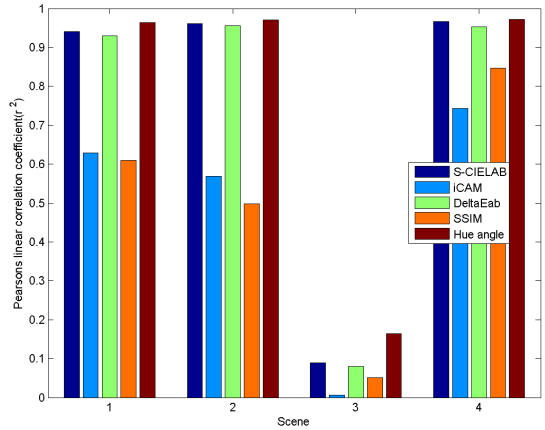


Figure 15. Whole image r^2 score for the different algorithms and scenes

same principles as ΔE_{ab}. SSIM should perform rather good due to the lightness changes, but this is not the case. iCAM shows a high difference but due to the confidence interval it cannot be rated worse than ΔE_{ab} or S-CIELAB. For the JP scene all metrics have problems with the regions. This shows that the metrics have a problem for images were small regions are altered, and the grade of visibility of these alterations.

Figure 15 shows the correlation for the different algorithms over the 4 scenes. The hue angle algorithm perform well in scene 1 (Girl),2 (Tore) and 4 (Cartoon), and in scene 3 (JP) this algorithm encounter problems along with all the other algorithms. This is because of the high visibility of errors in central parts of the image. SSIM and iCAM generally perform worse than the other algorithms.

4.3 Image difference metrics applied to eye tracker regions

In this part the frequency maps from the eye tracker have been applied to the image difference metrics. This map was based on the frequency map from the eye tracker,[17] which was calculated as

$$F(x,y) = \left(\frac{T(x,y)}{N(x,y)}\right)_{Nol} \tag{1}$$

where

$$F(x,y) \in [0,1]$$

The total time (T) an observer used on each pixel was divided by the number of times (N) the observer fixated on that pixel. The result was then normalized (Nol) by the maximum value in the map, and a gaussian filter was applied to the map to even out differences[17,22,23] and to simulate that we look at an area rather than one particular pixel.[24] The values from the metrics were multiplied with the frequency map to get the overall image difference. An example of a frequency map for one observers is shown in Figure 16.

$$MetricValue = \frac{\sum_x \sum_y (Mappixel_{x,y} \cdot Metricmap_{x,y})}{N_{numberofpixels}} \tag{2}$$

Some observers were discarded due to calibration or technical problems, for the different scenes 24 to 16 observers were used to create the frequency maps.

4.3.1 S-CIELAB

In S-CIELAB the points are located around the regression line. The Stripe images from the JP scene is the points with the biggest difference from the regression line, this also results in a low r^2 score of only 0.16 for S-CIELAB in this scene (Table 2). The three other scenes score between 0.92 and 0.97, this indicates a strong correlation. In the overall score (Figure 17) we can see that the LP5 and LM5 in the Cartoon and Girl scene get different scores even though they are rated similar by the observers.

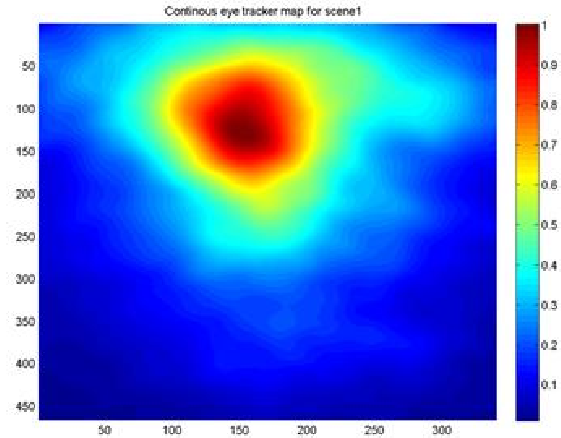


Figure 16. Continuous eye tracker map for scene 1

Scene/Algorithm	S-CIELAB	iCAM	ΔE_{ab}	SSIM
1 (Girl)	0.92	0.64	0.93	0.66
2 (Tore)	0.92	0.73	0.93	0.69
3 (JP)	0.16	0.08	0.12	0.09
4 (Cartoon)	0.97	0.73	0.96	0.80

Table 2. r^2 value for the different algorithms, eye tracker map applied

4.3.2 iCAM

iCAM gets an overall r^2 score of 0.37, and the points are scattered as seen in Figure 18. The Cartoon scores best of the scenes in iCAM with 0.73 with the Tore scene almost similar. The biggest problem for iCAM is the LM5 images, iCAM underestimates the score and rates the LM5 as the LM3 and LP3. The same happens in the Girl scene, and results in a r^2 of 0.64. For the JP scene none of the points are located near the regression line, resulting in a low correlation.

4.3.3 ΔE_{ab}

ΔE_{ab} gets an overall score of 0.65. Some points are located on the regression line as seen in Figure 19, but the stripe2 images are far off the regression line. We do also see that some images are given approximately the same score but the z-score difference is almost 0.5. For the Girl, Tore and Cartoon have a correlation of 0.93, 0.93 and 0.96. This indicate that ΔE_{ab} has a good correlation between the z-score and the predicted image difference. The reason for the low overall score is mainly from the JP scene.

4.3.4 SSIM

The overall r^2 value for SSIM is close to 0, this indicate almost no correlation between the points and the regression line, but as discussed earlier this is not a valid result. For the single scenes SSIM shows some correlation. In the Girl scene the r^2 value is 0.66 and Tore scores almost the same. The JP gives a score of almost 0.1, but most of the points are located more than 0.2 from the regression line (Figure 20). Images with more than 0.8 difference in z-score have been rated almost the same. Cartoon is the scene that scores the best for SSIM, with a r^2 of 0.80. Here most of the points are located close to the regression line.

4.3.5 Overall score for metrics

We can see from Figure 21 that S-CIELAB and ΔE_{ab} scores approximately the same in each scene, while iCAM and SSIM are more similar in their r^2 score. For the JP scene (scene 3) all metrics score low, but S-CIELAB has the highest value while iCAM has the lowest.

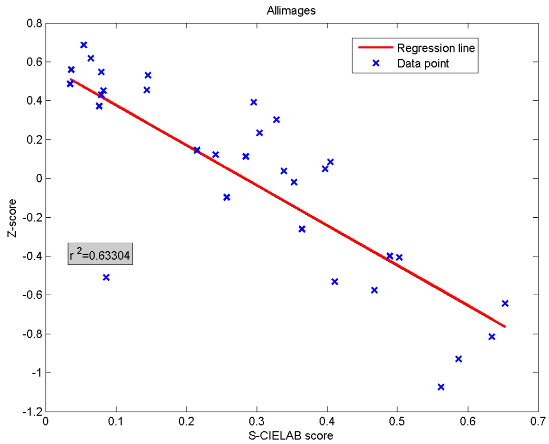


Figure 17. Regression plot for S-CIELAB weighted with eye tracker map

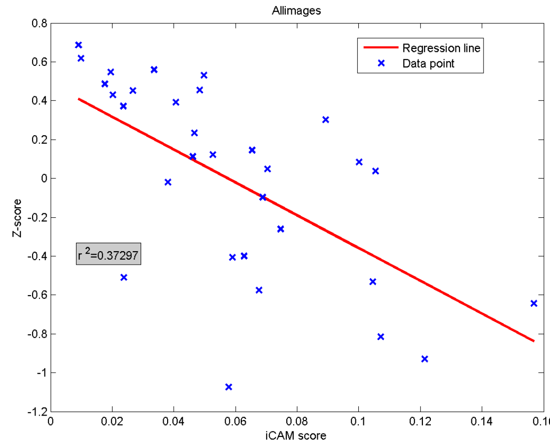


Figure 18. Regression plot for iCAM weighted with eye tracker map

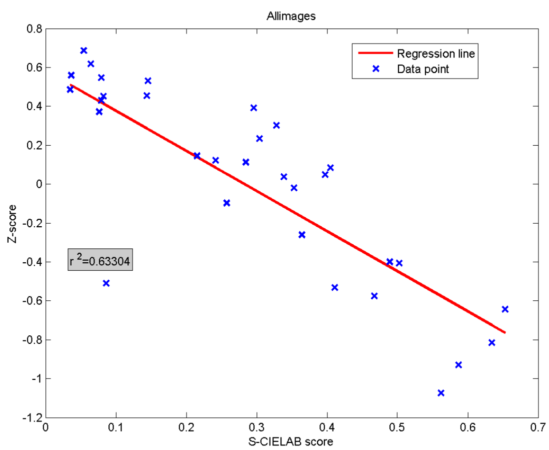


Figure 19. Regression plot for ΔE_{ab} weighted with eye tracker map

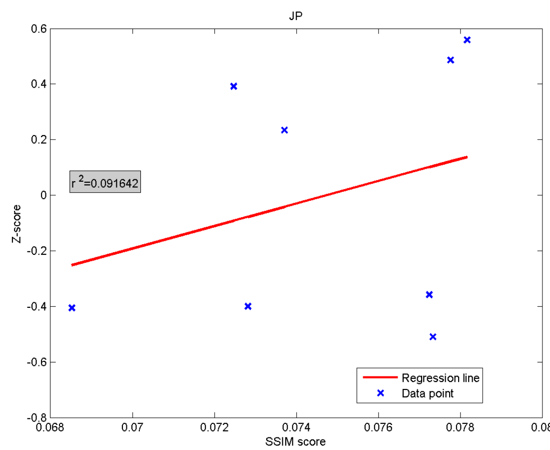


Figure 20. Regression line plot SSIM for JP data points

If we look at the mean squared difference from the regression line (Figure 22), S-CIELAB and ΔE_{ab} have almost the same mean. iCAM and SSIM has a higher mean value, but iCAM cannot be differeniated from the rest. S-CIELAB can be rated better than SSIM.

From this S-CIELAB and ΔE_{ab} seems to be the algorithms when we take into account the correlation coefficient.

4.4 Difference between the two methods

The results from the image difference metrics (Figure 15) is compared to the results from the eye tracker weighted results (Figure 21). In the first scene the two ways score almost the same corralation as seen in Figure 23, but SSIM improves slightly with 0.05. In the second scene SSIM and iCAM get a higher correlation, this is mainly because of the improvement in the LP5 and LM5 images in the Tore scene. In the third scene all metrics have an increased correlation coefficient. This third scene (JP) has regions that are small and highly visible even though the overall change is small. The eye tracker map has a higher weighting in this area, and this results in an increased correlation. In the fourth scene S-CIELAB, iCAM and ΔE_{ab} show only minor differences, while SSIM shows a result of almost -0.05.

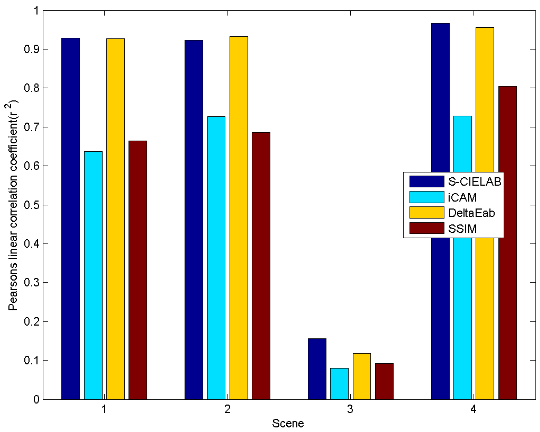


Figure 21. Correlation for the metric for eye tracker

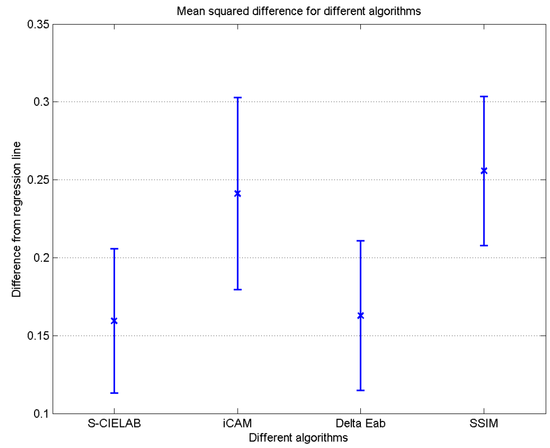


Figure 22. Mean squared difference - eye tracker

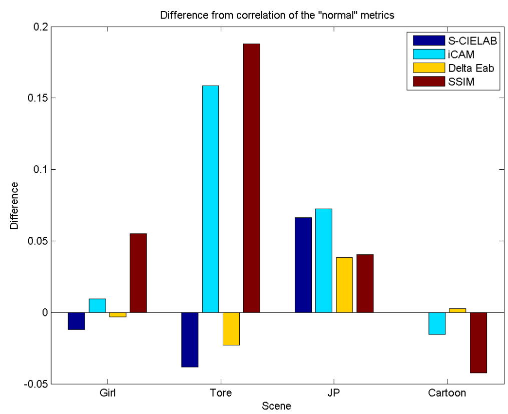


Figure 23. Difference from the normal and the eye tracker weighted map

5. CONCLUSION

We have shown that applying eye tracker frequency maps to image difference metrics can improve the overall correlation between the perceived image difference and the predicted image difference. Metrics that perform well before applying the eye tracker map only have minor changes, but metrics that perform moderate increase their correlation. Also in scenes where small regions are altered with highly visible changes, applying an eye tracker map will improve the performance of the metric.

6. FURTHER RESEARCH

Using an eye tracker each time improve image difference metrics is time consuming and not very practical. Further research would be to implement and test automatic gaze simulating region-of-interest algorithms, with the results given here this could eliminate the need for using an eye tracker to create the frequency maps.

REFERENCES

1. Eriko Bando, Jon Y. Hardeberg, and David Connah. Can gamut mapping quality be predicted by color image difference formulae. In *Human Vision and Electronic Imaging X, ed. B. Rogowitz, T. Pappas, S. Daly, Proc. of SPIE - IST Electronic Imaging,SPIE*, volume 5666, pages 180 – 191, 2005.

2. Nicolas Bonnier, Francis Schmitt, Hans Brettel, and Stephane Berche. Evaluation of spatial gamut mapping algorithms. In *Proceedings of IS&T and SID's 14th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, volume 14, pages 56–61, November 2006. ISBN / ISSN: 0-89208-292-5.
3. Guowei Hong and M. Ronnier Luo. Perceptually based colour difference for complex images. In Allan Rodrigues Robert Chung, editor, *Proceedings of SPIE: 9th Congress of the International Colour Association*, volume 4421, pages 618–621, 2002.
4. Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
5. Xuemei Zhang and Brian A. Wandell. A spatial extension of CIELAB for digital color image reproduction. *Proc. Soc. Inform. Display 96 Digest*, pages 731–734, 1996.
6. Xuemei Zhang, D. Amnon Silverstein, Joyce E. Farrell, and Brian A. Wandell. Color image quality metric S-CIELAB and its application on halftone texture visibility. In *COMPCON97 Digest of Papers*, pages 44–48, 1997.
7. Maria Vanrell, Ramon Baldrich, Anna Salvatella, Robert Benavente, and Francesc Tous. Induction operators for a computational colour-texture representation. *Computer Vision and Image Understanding*, 94(1-3):92–114, April-June 2004.
8. Jiangtao Kuang, Garrett M. Johnson, and Mark D. Fairchild. iCAM for high-dynamic-range image rendering. In *APGV '06: Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 151–151, New York, NY, USA, 2006. ACM Press.
9. Fritz Ebner and Mark D. Fairchild. Development and testing of a color space (IPT) with improved hue uniformity. In *IS&T/SID Sixth Color Imaging Conference: Color Science, Systems and Applications*, pages 8–13, 1998. ISBN / ISSN: 0-89208-213-5.
10. Garrett M. Johnson. *Measuring images: differences, quality and apperance*. PhD thesis, Rochester Institute of Technology, March 2003.
11. Mark D. Fairchild. Considering the surround in device-independent color imaging. *Color Research and Application*, 20:352–363, 1995.
12. Mark D. Fairchild. Modeling color appearance, spatial vision, and image quality imaging. *Color Image Science: Exploiting Digital Media*, pages 357–370, 2002.
13. Garrett M. Johnson and Mark D. Fairchild. A top down description of S-CIELAB and CIEDE2000. *Color Research and Application*, 28(6):425–435, December 2003.
14. Nathan Moroney, Mark D. Fairchild, Robert W.G. Hunt, Changjun Li, M. Ronnier Luo, and Todd Newman. The CIECAM02 color appearance model. *IST/SID 10th Color Imaging Conference*, pages 23–27, 2002.
15. Pei-Li Sun and Jan Morovic. What differences do observers see in colour image reproduction experiments. In *CGIV 2002: The First European Conference on Colour Graphics, Imaging and Vision*, 2002.
16. Jan Morovic and Pei-Li Sun. Visual differences in colour reproduction and their colorimetric correlates. *IS&T Tenth Color Imaging Conference*, Scottsdale, Arizona:292–297, 2002.
17. Jie Bai, Toshiya Nakaguchi, Norimichi Tsumura, and Yoichi Miyake. Evaluation of image corrected by retinex method based on S-CIELAB and gazing information. *IEICE trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol.E89-A(11):2955–2961, 2006.
18. Dennis R. Ankrum. Viewing angle and distance in computer workstations. 13/11/06: http://www.combo.com/ergo/vangle.htm.
19. SMI (SensoMotoric Instruments). 02/04/07: http://www.smi.de.
20. CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical Report ISBN: 3 901 906 26 6, CIE TC8-08, 156:2004. ISBN: 3 901 906 26 6.
21. Jan Morovic. *To Develop a Universal Gamut Mapping Algorithm*. PhD thesis, University of Derby, 1998.
22. Jason S. Babcock. Eye tracking observers during color image evaluation tasks. Master's thesis, Rochester Institute of Techonology, 2000.
23. David S. Wooding. Fixation maps: quantifying eye-movement traces. In *ETRA '02: Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 31–36, New York, NY, USA, 2002.
24. John M. Henderson, Carrick C. Williams, Monica S. Castelhano, and Richard J. Falk. Eye movements and picture processing during recognition. *Perception & Psychophysics*, 65:725–734, 2003.