Task-Based Accessibility Measurement of Daltonization Algorithms for Information Graphics

Joschua Thomas SIMON-LIEDTKE, Jon Yngve HARDEBERG
The Norwegian Colour and Video Computing Laboratory, Gjøvik University College

ABSTRACT

Color deficient people make up about eight percent of the male population and they are often confronted with problems when retrieving information from color information graphics like transportation or geographic maps. So-called daltonization algorithms to improve images for color deficient people have been widely discussed, but it has been difficult to compare and analyze the different strategies with psycho-physical experiments due to the vast time consumption of such setups and the somewhat rarity of color deficient observers. Thus, we propose a framework that compares different algorithms based on a task-fulfilling experiment and the use of simulation algorithms in order to use normal sighted observers as "virtual" color deficient observers. We found out that both the accuracy and the variation of the reaction time can be used as an indicator for good or bad algorithms. We also related it to the color differences among the colors in the graphic and propose an objective measurement based on lightness and chroma as starting point for future measurement methods and daltonization algorithms.

1. INTRODUCTION

The nature of color deficiencies is relatively well understood (Rigden 1999), and well established models exist to simulate color deficiencies (Brettel *et al.* 1997, Kotera 2012, Viénot *et al.* 1999). Based on these models, so-called daltonization algorithms have been proposed by Anagnostopoulos *et al.* (2007), Kotera (2012) and others, which can adjust the color palette of images (and information graphics) so as to improve their quality and accessibility in terms of retrieving information content for color deficient people.

However, evaluating different daltonization algorithms is difficult due to the limited number of color deficient observers available for the researcher and due to the complexity and time consumption of traditional psycho-physical methods like rank order or pair comparison. Likewise, choosing a good color palette is becoming one important goal of making graphics more accessible in the light of universal design. The goal of this paper is to introduce a framework that can measure the effectiveness of daltonization algorithms specifically and the accessibility of information graphics generally. The paper is based on the premise that a well accessible image allows both color deficient and normal sighted observers to retrieve information quickly from the graphic. This is why we based our experiment on the reaction time needed for an observer to fulfill a certain task.

On the one hand, we decided to simulate the original and daltonized images such that normal sighted people can be used as "virtual" color deficient people assuming that the used algorithms simulate color deficiencies accurately. On the other hand, we chose a task-fulfilling experiment as setup for the framework in order to reduce experimentation time and in order to create a more life-like experience for the observer.

2. METHOD

To begin with, we compute the simulation, the daltonized version of the image and the simulated version of the daltonized image given a particular information graphic. For the original image, we identify the patches and colors of interest, i.e. the colors of the transportation lines that the observer will be asked to identify. For all of these patches we extract the RGB color values from all versions.

Secondly, we show randomly each of the different versions and the original to the observer on a gray background (c.Fig.1-lft). Each of the possible color patches associated with the particular image version are shown next to the target patch in a random order. The observer is then asked to click on the target patch in order to start the experiment, and then he/she has the task to click as quickly as possible on the right transportation line inside of the map. When the correct transportation line is clicked, the program proceeds to the next color patch. Also, the exact time needed to find the right transportation line is stored in a database, together with the exact position of the click inside the map and the information about whether or not the observer succeeds to click on the right line on the first try.

For our experimentation, we chose three different images of public transportation maps from the Norwegian cities Gjøvik, Bodø and Oslo. Also, we tested two different simulation and daltonizing algorithms based on Kotera (2012) and Viénot *et al.* (1997)/ Anagnostopoulos (2007). Thus, we get seven different versions per image - a total of 21 images for the experiment. With respect to the patches, we had five different transportation lines for the Bodø images, six for the Oslo images and seven for the Gjøvik images.

Moreover, we had a total of 23 observers, typically students between 20-30 years old; none of the observers were color deficient. The experiment was implemented as a website running on a development server on the local machine presented on a calibrated laptop screen. The website was presented in a Chrome browser window in full screen. The experiment was setup in a controlled environment with simulated D50 lighting.

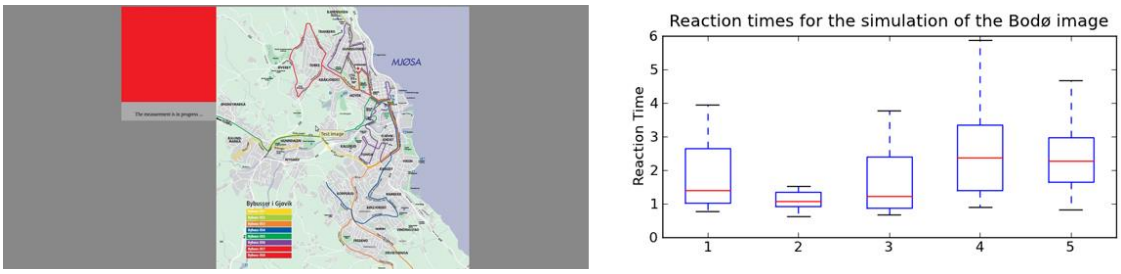

Figure 1: Website setup of our experiment (left) and reaction times (right) for the Anagnostopoulos simulation of the Bodø map

3 RESULTS AND DISCUSSION

In the following plots, we compared three different attributes of the images: To begin with, we plotted the reaction time for each version: On the one hand over all patches for each version (Fig.3-left) and on the other hand for each patch individually for each version (Fig.1-right). Secondly, we counted the number of correct clicks and divided them by the total number of clicks for all observers. Again, we plotted this accuracy rate summed up for all patches for each version (Fig.3-right) and for each patch individually for each version (Fig.2-right). Lastly, we computed the color difference between each individual patch in comparison to every other patch for each version for each image (Fig.2-left).

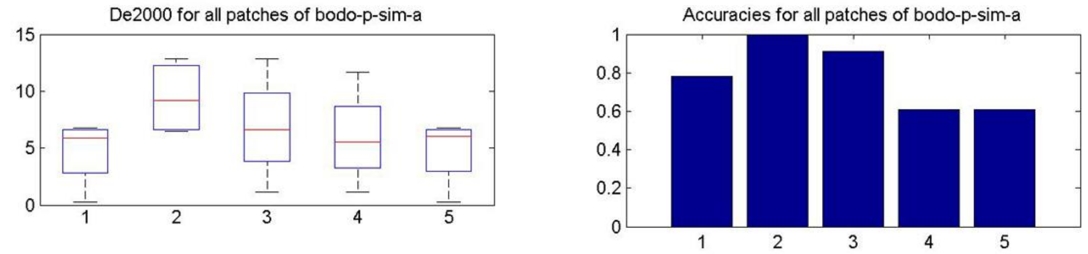

Figure 2: Color differences (left) and accuracies (right) for the Anagnostopoulos simulation of the Bodø map

First of all, we made the observation that although we cannot obtain any statistically significant difference from the reaction times, there is indeed a measurable tendency that a good daltonization reduces the reaction time of the observers and that easily accessible images have a lower variation among all observers (Fig.3-left).

Secondly, we can observe that the accuracies are high for easily accessible images and that they actually increase for the daltonized images and the simulated versions of the daltonized images (Fig.3-right). Also, it can be seen that for high accuracies, we also have fast reaction times and, more importantly, the variation among all observers is lower (Fig.2-right & Fig.1-right).

Thirdly, we can make the observation that there is no measurable correlation between average or minimum color difference and accuracy. A low minimum color difference <u>might</u> decrease accuracy. However, we can observe that the algorithm by Anagnostopoulos increases accuracy and decreases variation, whereas the Kotera algorithm does not seem to have a significant effect in improvement (Fig.3). The visual difference between both algorithms is that the Anagnostopoulos algorithm tries to maintain the hues of the original color and changes lightness and chroma, whereas Kotera changes colors globally, i.e. strong hue shifts exist. This observation might be a starting point for future accessibility metrics.

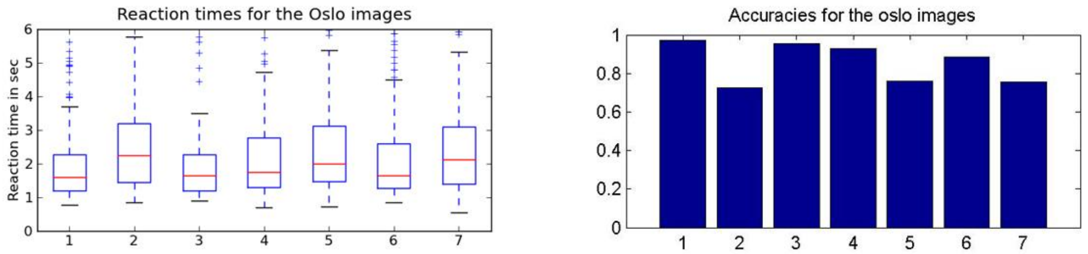

Figure 3: Overall reaction times (left) and overall accuracies (right) for all Oslo images[1]

Finally, it makes perfectly sense to use the setup in form of the task-fulfilling experiment because it saves a lot of time in comparison to psycho-physical experiments, in which observers have to compare and judge many pairs of images. In average, one

[1] 1-Original image, 2-Simulation (Anagnostopoulos), 3-Daltonization (Anagnostopoulos), 4-Simulation of daltonization (Anagnostopoulos), 5-Simulation (Kotera), 6-Daltonization (Kotera), 7-Simulation of daltonization (Kotera)

observer had to spend between five and ten minutes including the explanation per experiment. The general subjective feedbacks from the observers were positive as they found the task more interesting than for example other experimentation with solely pair-wise or rank-order comparison. Also, finding the right transportation line in an image corresponds very to a real-life scenario. In future research, the experiment should additionally be repeated with color deficient observers in order to compare their behavior with the “virtual” color deficient observers.

4. CONCLUSIONS

The accuracy on the one hand appears to be a suitable indicator for a good or bad performance of a given daltonization algorithm. Thus, we suggest changing the framework to focus stronger on the accuracy aspect of the task-fulfilling experiment. Nevertheless, the reaction times should still be recorded because the variation of the reaction times among observers gives a good indicator for the quality as well and the average reaction time gives a general tendency on how the algorithm performs. Moreover, we suggest building an objective image quality metric for color deficient people based on lightness and chroma rather than a metric build on color difference generally.

ACKNOWLEDGEMENTS

We want to thank Prof. Ivar Farup for mathematical and implementation support and Dr. Peter Nussbaum and Dr. Marius Pedersen for support and feedback.

REFERENCES

Anagnostopoulos, C.-N., G. Tsekouras, I. Anagnostopoulos, & C. Kalloniatis 2007. *Intelligent modification for the daltonization process of digitized paintings*. Proceedings of the 5th International Conference on Computer Vision Systems, Bielefeld University.

Brettel, H., F. Viénot & J.D. Mollon 1997. *Computerized simulation of color appearance for dichromats*. J. Opt. Soc. Am. A, 14 (10): 2647–2655.

Kotera, H. Nov. 2012. *Optimal daltonization by spectral shift for dichromatic vision*. Twentieth Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications. Los Angeles, California: 302–308.

Rigden, C. 1999. *“The eye of the beholder” - designing for colour-blind users*. British Telecommunications Engineering, 17 (4): 291–195.

Sharma, G., W. Wu & E. N. Dalal 2005. *The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations*. Color Research & Applications, 30 (1): 21–30.

Viénot, F., H. Brettel & J. D. Mollon 1999. *Digital video colourmaps for checking the legibility of displays by dichromats*. Color Research & Applications 24 (4): 243–252.

Address: Joschua Thomas Simon-Liedtke,
Gjøvik University College,
P.O.Box 191, N-2802 Gjøvik, Norway
E-mails: thomas.simon@hig.no, jon.hardeberg@hig.no