

Multiscale Daltonization in the Gradient Domain

Joschua Thomas Simon-Liedtke and Ivar Farup[^]

The Norwegian Colour and Visual Computing Laboratory, Department of Computer Science, NTNU—Norwegian University of Science and Technology, Gjøvik, Norway
E-mail: ivar.farup@ntnu.no

Abstract. We propose a daltonization method that enhances chromatic edges and contrast for color-deficient people by optimizing the gradient of an image. We rotate and scale the error gradient between the original and its simulation in the color space into the direction of optimal visibility that is orthogonal to both the main direction of information loss and the direction of lightness. Then, we reintegrate the daltonized image version from the modified gradient through an iterative diffusion process. Moreover, we include multiscaling to guarantee optimal daltonization on different scales of the image. Also, we provide an interface for data attachment modules designed to maintain naturalness of memory colors like neutral colors. We evaluate and compare our proposed method to other top-performing daltonization methods in behavioral and psychometric experiments. A visual-search experiment assessing performance of the attentional mechanism of the human visual system before and after daltonization measures the greatest improvement in accuracy for our proposed method compared to the original and all investigated daltonization methods. It also reveals optimal results for both natural and Ishihara images among both protan and deutan color-deficient observers. Furthermore, we can deduce from the results of a pairwise preference evaluation that our proposed method is preferred highest amongst all color-deficient people in total. Our proposed method is also ranked among the most preferred daltonization methods for both protan and deutan color-deficient observers individually. © 2018 Society for Imaging Science and Technology.
[DOI: 10.2352/J.Percept.Imaging.2018.1.1.010503]

1. INTRODUCTION

Trichromatic color vision (referred to as normal-sighted color vision) is initiated by photosensitive cones carrying three different types of pigments on the retina of the human eye [1]. However, around 8% of the male population have a hereditary color vision deficiency (CVD) [2, 3]. CVDs are most often caused by non-existent pigments of one of the cones (so-called *dichromacy*), or cones whose pigments have slightly shifted sensitivities as compared to normal-sighted observers (so-called *anomalous trichromacy*) [1]. We refer to CVDs related to L-cones as *protan* CVDs, whereas anomalies related to the M-cones are called *deutan* CVDs. Protan and deutan color-deficient people have difficulties in differentiating colors between red and green. More precisely, certain colors (so-called *confusion colors*) are easily distinguishable for the normal-sighted, at the same times as

they are hardly discriminable or perceived identically by the color-deficient [4, 5].

Furthermore, color-deficient vision is characterized by the decreased capability of detecting chromatic edges and contrast [1]. However, contour information described by edges and contrast plays a significant role in the human visual system (HVS) [1, 6]. More precisely, a reduced capability to detect chromatic edges and contrast might lead to behavioral disadvantages related to, for example, the attentional system by reducing the “pop-out” effect of certain colors [7, 8]. Other disadvantages might affect object recognition and the detection of emotional states as well [8–10]. Simulations of colour vision deficiencies are shown in Figure 1.

Therefore, daltonization methods exist to adapt confusion colors in digital images to increase their discriminability [11–14]: Kotera [15] introduced a recoloring method to preserve naturalness and enhance visibility for the color-deficient of pseudo-spectral “fundamental images” computed from spectral LMS space. Fidaner et al. [16] proposed a recoloring method optimized for protanopes in which they redistribute the error image for the color-deficient in RGB. Most of these daltonization methods change confusion colors globally. However, we argue that spatial approaches targeting local color changes might lead to promising results due to the importance of edges and contrast for the HVS [1].

We typically address edges and contrast in the gradient domain [17, 18]. In general, we can reintegrate images from a modified gradient following the general strategy for multidimensional differential equations by solving the Euler–Lagrange equation as in [19] (typically solved numerically as in [20]).

Last, the assessment of daltonization methods is important to verify their performances. We presented behavioral methodologies to assess the performance of daltonization on the attentional mechanism by using different visual-search experiments in [21–23]. We concluded from these studies that especially the accuracy data is well suited as an interval scale to rank different daltonization methods. Moreover, psychometric scaling experiments can complete the evaluation by assessing preference or naturalness of daltonized images by using, for example, pairwise comparison following Thurstone’s Law of Comparative Judgment [24, 25].

We propose a novel daltonization method, Yoshi-II, that focuses on enhancing chromatic edges and contrast. Yoshi-II is a content-dependent, spatial, customized, mainly

[^] IS&T Member

Received June 11, 2018; accepted for publication July 20, 2018; published online Nov. 15, 2018. Associate Editor: Michael A. Webster.
2575-8144/2018/1(1)/010503/12/\$00.00

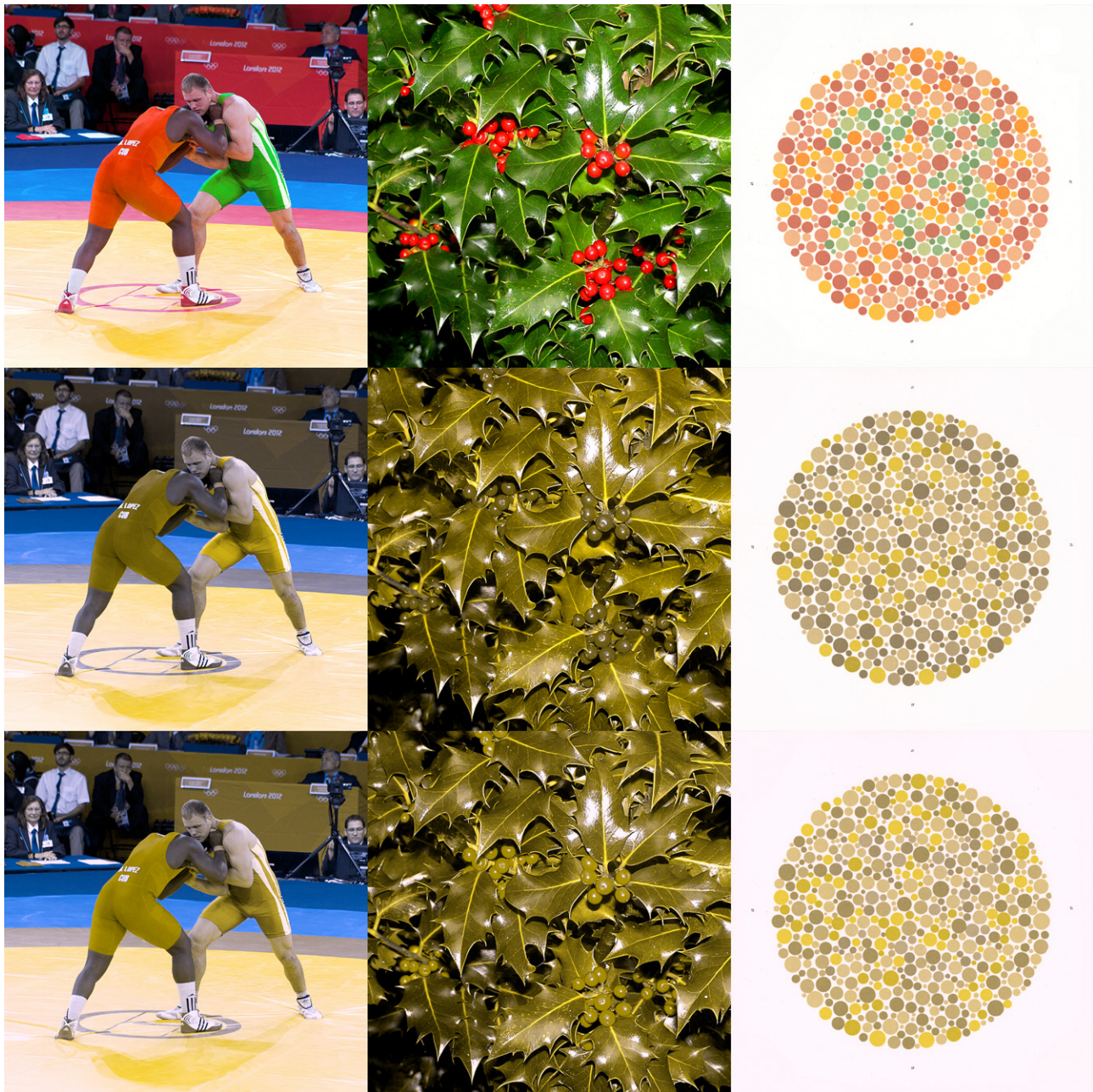


Figure 1. The first row shows the original images. The second row shows the protan simulations and the third row shows the deutan simulations. The images have been simulated using the Brettel simulation for dichromats [26].

automated (with possibilities for user input) recoloring method according to the classification in [11]. Computations for Yoshi-II take place in the gradient domain. We also use computations in scale space to process chromatic edges and contrast on different scales of the image, and provide a data attachment link to preserve naturalness of memory colors.

We summarize Yoshi-II in the following steps:

- (i) We first obtain an error image between the original and its CVD simulation in gradient space.
- (ii) Second, we modify the original gradient by rotating and scaling the error image in the color space into the direction of optimal visibility that is orthogonal

to the direction of main error and the direction of lightness. Then, we add the modified gradient to the original gradient. This modification preserves the general contrast of the image and reintroduces lost information.

- (iii) We eventually obtain the daltonized image from the modified gradient image by solving the Euler–Lagrange equation and applying the Poisson diffusion.

Moreover, computations in scale space guarantee optimal results for confusion colors on different resolutions scales and areas of confusion colors that are not directly adjacent to each other [27]. We also introduce an interface

for data attachment modules to maintain the naturalness of memory colors. We compare our proposed method with other state-of-the-art daltonization methods using behavioral and psychometric experimentations. Finally, we discuss possible improvements for our proposed method.

It should be noted that the daltonization method presented in this paper is specific for individual observer types and not meant as a general improvement for everybody. A general enhancement that will work for normal observers, as well as for observers with CVD is not goal of this paper. A discussion about daltonization in general, and characteristics and requirements of a “good” daltonization method can be found in a previous article [28].

2. PROPOSED DALTONIZATION ALGORITHM

An image is represented as a three-dimensional function \mathbf{u}_0 with typically three channels, red, green and blue:

$$\mathbf{u}_0 : \Omega \rightarrow c, \quad \text{where } \Omega \subset \mathbb{R}^2, \quad \text{and } c = [0, 1]^3 \quad (1)$$

$$\mathbf{u}_0 = \begin{bmatrix} \mathbf{u}_R & \mathbf{u}_G & \mathbf{u}_B \end{bmatrix}.$$

The gradient $\nabla \mathbf{u}_0$ represents the change of image values along the x - and y -axis for all channels [29]:

$$\nabla \mathbf{u}_0 : \Omega \rightarrow \mathbb{R}^2 \times \mathbb{R}^3 \quad (2)$$

$$\nabla \mathbf{u}_0 = \begin{bmatrix} \frac{\partial \mathbf{u}_0}{\partial x} \\ \frac{\partial \mathbf{u}_0}{\partial y} \end{bmatrix}.$$

The simulated image $\mathbf{s}(\mathbf{u}_0)$, i.e., the representation of color-deficient vision, and its gradient image $\nabla \mathbf{s}(\mathbf{u}_0)$ are computed by using any simulation method [5, 15, 26, 30, 31]. We decided to use the Brettel simulation method in the proposed implementation, because it represents one of the most accurate simulation method as we discussed in a previous paper [32].

First, we compute the difference image \mathbf{u}_0 between the original image and its simulation: $\mathbf{d}_0 = \mathbf{u}_0 - \mathbf{s}(\mathbf{u}_0)$. \mathbf{d}_0 contains the information that is lost for color-deficient observers. We apply a principal component analysis (PCA) [33] on \mathbf{d}_0 to obtain its first principal component, normalized and denoted as \mathbf{e}_d , that represents the main direction of difference. Colors or color changes along \mathbf{e}_d are least visible for color-deficient observers. It is the direction, along which most information gets lost for the color-deficient.

Second, we compute the direction of optimal visibility \mathbf{e}_c that is orthogonal to both the direction of maximum lost \mathbf{e}_d and the direction of lightness \mathbf{e}_l to preserve the lost information. \mathbf{e}_c (the c stands for chroma) is computed by:

$$\mathbf{e}_c = \mathbf{e}_d \times \mathbf{e}_l. \quad (3)$$

Orthogonality to the error vector guarantees minimal information loss. However, rotation along only this direction would merely change the lightness of the confusion colors while maintaining hue and chroma. In [34], we tried to include lost information only in the lightness channel with

unsatisfactory results. Thus, we concluded that the rotation should also be orthogonal to the direction of lightness. The lightness vector can, on the one hand, be defined by the neutral colors in sRGB space: $\mathbf{e}_l = [1.0, 1.0, 1.0]$. This choice makes sense because neutral colors are perceived identically by color-deficient and normal-sighted people [26]. The lightness vector could, on the other hand, also be defined as the direction that is orthogonal to the plane of constant lightness in sRGB space: $\mathbf{e}_l = [0.2126, 0.7152, 0.0722]$ [35]. This direction would change hue and chroma of confusion colors while maintaining their original lightness.

Third, we approximate the difference between the original and its simulation, i.e., the lost information, through its projection along \mathbf{e}_d , $\nabla \mathbf{u}_0 \cdot \mathbf{e}_d$, rotate and scale the lost information into the direction of optimal visibility \mathbf{e}_c , and add the rotated and scaled information to the original gradient to obtain a modified gradient denoted as tensor \mathbf{G} :

$$\mathbf{G} = \nabla \mathbf{u}_0 + (\nabla \mathbf{u}_0 \cdot \mathbf{e}_d)(\chi \mathbf{e}_c) \quad (4)$$

χ influences how much the chroma should change at any point.

Fourth, we obtain the optimal value for χ at each point by assuming that the norm of the simulated tensor $\mathbf{S}(\mathbf{G})$ is equal to the norm of the original gradient because the visible gradient for a color-deficient observer looking at the daltonized image should equal in magnitude the gradients for a normal-sighted observer looking at the original image:

$$\|\mathbf{S}(\mathbf{G})\|_F^2 \stackrel{!}{=} \|\nabla \mathbf{u}_0\|_F^2 \quad (5)$$

where $\|\dots\|_F^2$ represents the Frobenius norm. In this step, we approximate the simulation method by a linear projection onto the $\mathbf{e}_d, \mathbf{e}_l$ plane. Consequently, the actual daltonization element of the tensor, e.g., $\chi (\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c$, is unaffected by the approximated simulation because it represents the part that is orthogonal to this projection. Thus, the simulated tensor can be approximated by $\tilde{\mathbf{S}}(\mathbf{G}) \approx \nabla \mathbf{s}(\mathbf{u}_0) + \chi (\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c$. However, $\tilde{\mathbf{S}}(\mathbf{G}) \neq \mathbf{S}(\mathbf{G})$. Equation (5) can be developed as:

$$\begin{aligned} \|\tilde{\mathbf{S}}(\mathbf{G})\|_F^2 &\stackrel{!}{=} \|\nabla \mathbf{u}_0\|_F^2 \\ \|\nabla \mathbf{s}(\mathbf{u}_0) + \chi (\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c\|_F^2 &\stackrel{!}{=} \|\nabla \mathbf{u}_0\|_F^2 \\ \Rightarrow \chi^2 \|(\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c\|_F^2 + 2\chi (\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c : \nabla \mathbf{s}(\mathbf{u}_0) & \\ + \|\nabla \mathbf{s}(\mathbf{u}_0)\|_F^2 - \|\nabla \mathbf{u}_0\|_F^2 &= 0. \end{aligned} \quad (6)$$

Equation (6) is a quadratic equation for χ solved by:

$$\chi_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (7)$$

where $a = \|(\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c\|_F^2$, $b = 2(\nabla \mathbf{u}_0 \cdot \mathbf{e}_d) \mathbf{e}_c : \nabla \mathbf{s}(\mathbf{u}_0)$, and $c = \|\nabla \mathbf{s}(\mathbf{u}_0)\|_F^2 - \|\nabla \mathbf{u}_0\|_F^2$. However, which sign should we choose for χ ? We integrate over the image for all the positive and all the negative χ and compare both values. If the positive integral is smaller than negative, we choose the positive χ values and vice versa. In other words, we want the solution that causes the least absolute changes in the image. We always

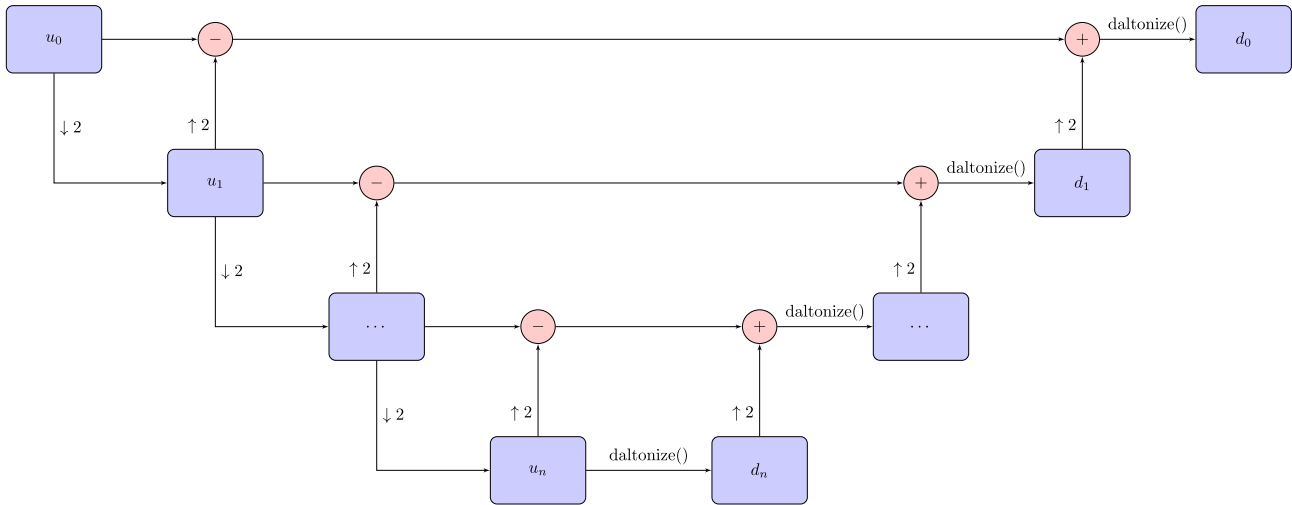


Figure 2. Multiscaling pyramid structure, where “ $\downarrow 2$ ” indicates minimizing the image to half its size, “ $\uparrow 2$ ” indicates enlarging it to double its size, and “*daltonize()*” indicates the daltonization step at each iteration.

choose the negative or positive χ value at each point in the image to change colors only toward one direction.

Fifth, reintegrating the improved daltonized image from the tensor \mathbf{G} can be formulated as variational problem, where we find the daltonized image \mathbf{u} through optimizing the following equation based on the Tikhonov model [36]. Ω represents the image domain, whereas dA represents $dx dy$ in image coordinates:

$$\min \frac{1}{2} \int_{\Omega} \|\nabla \mathbf{u} - \mathbf{G}\|_F^2 dA. \quad (8)$$

Solving the Euler–Lagrange equation [19, 37] results in a Poisson equation form:

$$\nabla^2 \mathbf{u} - \nabla \cdot \mathbf{G} = 0. \quad (9)$$

In the discrete case, this Poisson equation can be solved by gradient descent, i.e., the iterative reintegration by time [38]:

$$\frac{\partial \mathbf{u}}{\partial t} = \nabla^2 \mathbf{u} - \nabla \cdot \mathbf{G}. \quad (10)$$

We define a stop criterion c for the gradient descent based on the residuals between the updated versions and the tensor \mathbf{G} :

$$c = \frac{\sqrt{\int_{\Omega} \|\nabla \mathbf{u}_i - \mathbf{G}\|_F^2 dA} - \sqrt{\int_{\Omega} \|\nabla \mathbf{u}_{i+1} - \mathbf{G}\|_F^2 dA}}{\sqrt{\int_{\Omega} \|\nabla \mathbf{u}_i - \mathbf{G}\|_F^2 dA}}. \quad (11)$$

The gradient descent stops when c is smaller than a certain threshold indicating that changes in the updating process are not significant anymore from one iteration step to the next one.

Sixth and finally, we test four different strategies for the boundary conditions:

- (i) In the first strategy, based on the Dirichlet condition [39], we keep the boundary values of the updated image fixed to the values of the original image, i.e., $\mathbf{g}(x, y) = \mathbf{u}_0(x, y)$.

- (ii) Based on the Neumann condition [39], the boundary values of the gradient image are set to the gradient of the original image, i.e., $\mathbf{h}(x, y) = \nabla \mathbf{u}_0$.
- (iii) Another strategy based on the Neumann condition is to set the gradient image at the boundaries to zero, i.e., $\mathbf{h}(x, y) = 0$.
- (iv) Last, zero was assigned to the Laplacian $\nabla^2 \mathbf{u}(x, y)$ at the boundaries.

In the current implementation, we decided to use the third option because it gave the least amount of visual artifacts. Option (i) would create a visible contrast between original border pixels and daltonized center pixels, option (ii) would cause “bleeding” of unnatural colors from the borders, and option (iii) would cause “bleeding” of black from the borders.

One limitation of gradient images is that only areas of confusion colors that are directly adjacent to each other can be daltonized. One solution for this problem is multiscaling, e.g., daltonization of the image at different resolutions in scale space [27]. We can represent the core idea of multiscaling as a pyramid structure, which “is obtained by successively reducing the image size by combined smoothing and subsampling” [27]. In Yoshi-II (cf. Figure 2), we reduce the size of the image repeatedly by half until the lowest resolution, from where we start the actual daltonization. From then on, we daltonize the current scale version and add the result to the difference between the two previous resolution scales until we reach the highest resolution again. We can use different interpolation options for the re-sizing step including bilinear, bicubic, etc., interpolation, whereof we used bicubic interpolation in the current implementation because it caused the least amount of visual artifacts.

Another important requirement for daltonization methods is to maintain the natural appearance of the original image. Especially memory colors like skin, grass, etc., or neutral areas should stay the same as the original as much as possible. This can be done by adding a data attachment link

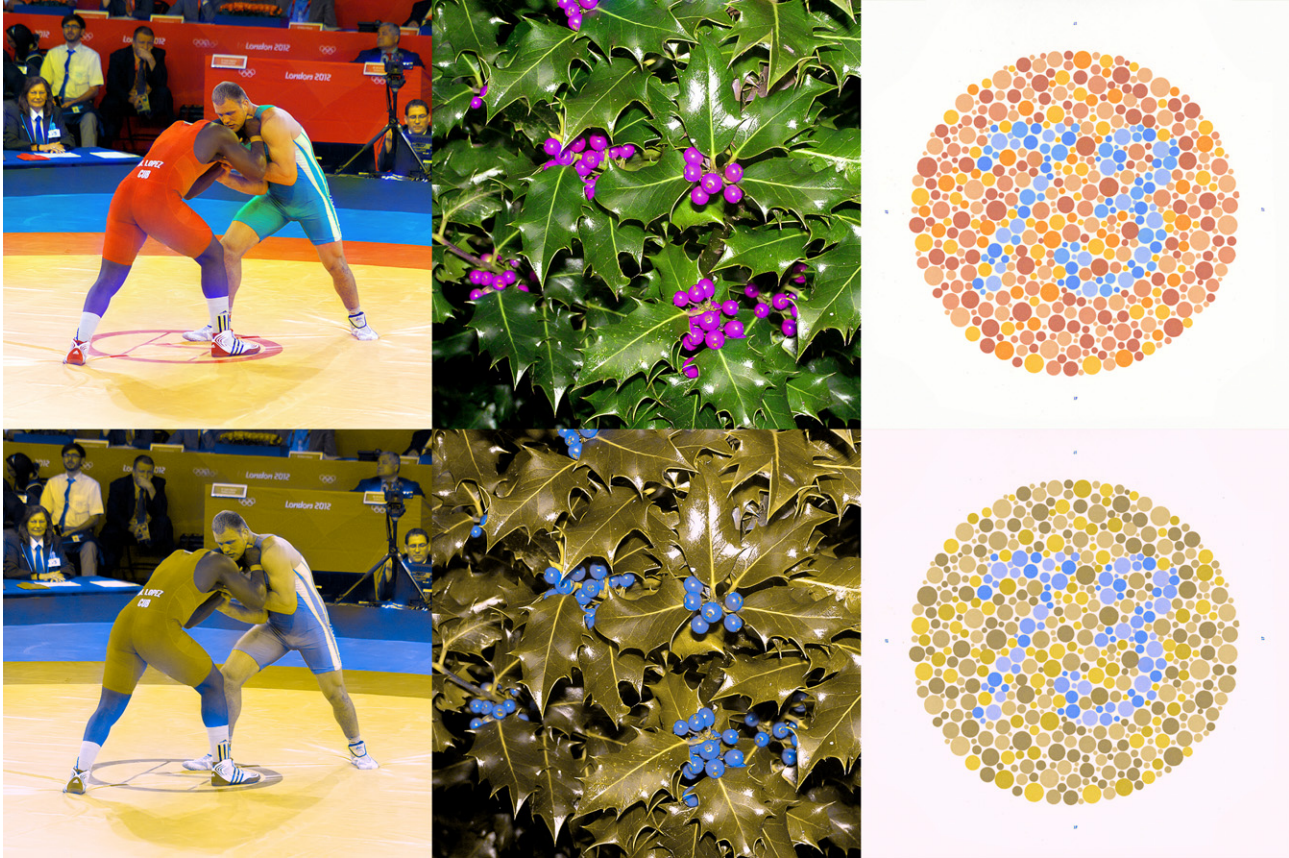


Figure 3. The first row shows the deutan daltonization by our proposed method Yoshi-II and the second row their simulations using the Brettel simulation for dichromats [26].

or fidelity term [36] to Equation (10) weighted by a function $h(x, y)$ and the strength of the data attachment indicated by the scalar l :

$$\frac{\partial \mathbf{u}}{\partial t} = [\nabla^2 \mathbf{u} - \nabla \cdot \mathbf{G}] - l \cdot h(x, y)(\mathbf{u} - \mathbf{u}_0). \quad (12)$$

In the current implementation of Yoshi-II, a data attachment is added to preserve neutral areas by using a weighting function $h(x, y)$ analyzing chroma. Chroma is computed in CIELAB color space as $C_{ab}^*(x, y) = \sqrt{a^*(x, y)^2 + b^*(x, y)^2}$ with $a^*(x, y)$ and $b^*(x, y)$ being the a^* and b^* coordinates in CIELAB space at (x, y) [40]. Moreover, the chroma is weighted by a Gaussian:

$$\begin{aligned} h(x, y) &= \exp\left(-\frac{C_{ab}^*(x, y)^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{a^*(x, y)^2 + b^*(x, y)^2}{2\sigma^2}\right). \end{aligned} \quad (13)$$

3. EVALUATION METHOD

For the experimental validation, we use our proposed daltonization method, Yoshi-II, with data attachment for neutral colors and multiscale. Simulated images are computed using the simulation method by Brettel et al. [26].

Re-sizing in scale space is done by using the `scipy.misc.resize` function with the `bicubic` interpolation option [41], and we set the cutoff threshold for the multiscale step to 0.00005. Moreover, we set the σ of the Gaussian in the neutral color fidelity term to 0.05. We use the constant lightness vector as defined in the sRGB standard. The sign for the quadratic solution of the χ values has been chosen in the very first daltonization step at the lowest resolution. That means if the coarsest resolution revealed the least difference for the positive solutions of the χ values, the positive signs have been chosen for all the following scale levels as well to provide consistent changes in recoloring. Last, we use the third boundary option described before. Examples for three images and their respective simulations for deutan and protan dichromatic vision are shown in Figure 3 and Figure 4, respectively. The examples provide a general idea about how the proposed method performs. It also showcases some of the advantages and disadvantages of Yoshi-II.

Each observer is tested and categorized using the HRR CVD test by Richmond Products, Inc. [42, 43]. We followed a guideline [44] suggesting a minimum of 10–20 observers for psychophysical and psychometric evaluations and recruited 13 normal-sighted, 14 deutan color-deficient, and 5 protan color-deficient observers for the behavioral experiment:



Figure 4. The first row shows the protan daltonization by our proposed method Yoshi-II and the second row their simulations using the Brettel simulation for dichromats [26].

- (i) Two of the deutan color-deficient observers had mild, three have medium, and nine have strong deutan CVDs.
- (ii) All of the five protan color-deficient observers had strong protan CVD.

This distribution roughly represents the real-life distribution of CVD that is reported with approximately 75% for deutan defects and 25% for protan defects (About 6% of the male population have a deutan defect versus 2% with a protan defect according to [2] or 6.09% versus 1.92% according to [3]). Moreover, the results from normal-sighted observers individually, deutan color-deficient observers individually, and protan and deutan color-deficient observers combined will be representative enough. The results from only protan color-deficient observers, however, will be taken with care. We also recruited 13 normal-sighted, 13 deutan color-deficient, and four protan color-deficient observers for the psychometric experiment.

- (i) Two of the deutan color-deficient observers had mild, two had medium, and nine had strong deutan CVD.
- (ii) All four protan color-deficient observers had strong protan CVD.

3.1 Behavioral Experiment

We use a behavioral method for the assessment called ViSDEM that we presented in [22, 23]. In this visual-search experiment, we measure the performance of the attentional system. The observer has to retrieve information related to color from various motives before and after daltonization. We compute the accuracies and response times of all methods from the observations of all observers. We rank the investigated methods on an interval scale from highest to lowest behavioral performance by their accuracy values in decreasing order. We compare our proposed method to the two best performing daltonization methods from [23]: Namely the Kotera [15] and the Fidaner methods [16]. We use 11 sets with 11 associated tasks containing 17 motives, of which 11 depicted natural images and six represented Ishihara plates. We only present the results from the accuracy data because the response time data does not give any additional insights as we discussed in [23]. We analyze the data with the same methods described in [23] including the Wilson interval score [45] and a χ^2 test [46].

3.2 Psychometric Experiment

The psychometric evaluation is based on Thurstone's Law of Comparative Judgment [24, 25]. More precisely, we follow the Case V solution described in [25]. We show each possible

Table I. Statistical analysis of the accuracy data for color-deficient versus normal-sighted (top), protan versus deutan (middle), and natural versus Ishihara images for color-deficient observers (bottom). Statistically significant values are emphasized.

	Fidaner	Kotera	Yoshi-ii		Fidaner	Kotera	Yoshi-ii
(a) P-values of the χ^2 test for color-deficient observers				(b) P-values of the χ^2 test for normal-sighted observers			
Original	8.6×10^{-5}	1.0×10^{-13}	2.7×10^{-29}	Original	0.52	0.03	0.02
Fidaner	x	3.1×10^{-4}	1.7×10^{-14}	Fidaner	x	6.6×10^{-3}	4.1×10^{-3}
Kotera	x	x	2.0×10^{-5}	Kotera	x	x	0.86
(c) P-values of the χ^2 test for deutan color-deficient observers				(d) P-values of the χ^2 test for protan color-deficient observers			
Original	0.11	4.3×10^{-14}	4.0×10^{-20}	Original	3.6×10^{-6}	8.0×10^{-3}	4.6×10^{-11}
Fidaner	x	1.2×10^{-9}	6.7×10^{-15}	Fidaner	x	0.04	0.03
Kotera	x	x	0.05	Kotera	x	x	3.9×10^{-5}
(e) P-values of the χ^2 test of natural images				(f) P-values of the χ^2 test of Ishihara images			
Original	0.21	0.22	3.2×10^{-3}	Original	1.9×10^{-6}	1.3×10^{-31}	5.7×10^{-38}
Fidaner	x	0.01	0.08	Fidaner	x	1.3×10^{-13}	2×10^{-18}
Kotera	x	x	3.6×10^{-5}	Kotera	x	x	0.10

pair between the daltonized and original versions of one motive. We ask the observers which one of the two versions in each observation “he/she prefers.” We sum the results from each comparison over all observers in the so-called observation matrix and compute the frequency/probability matrix by dividing it by the total number of observations. We then compute the inverse of the cumulative density function for each of the probability values to obtain the individual z-scores. The average of all z-scores for each version results in the mean z-score or “worth” value of the version. A higher z-score indicates that the observers were more likely to choose this version over another. Finally, we compute the Montag confidence intervals (CIs) of each z-score [47]:

$$CI = z \pm b_1(n - b_2)^{b_3}(N - b_4)^{b_5} \quad (14)$$

where $b_1 = 1.76$, $b_2 = -3.08$, $b_3 = -0.613$, $b_4 = 2.55$, $b_5 = -0.491$, N is the number of observations, and n is the number of investigated daltonization methods. Z-scores are statistically significantly different, when their CIs do not overlap. As before, we compare Yoshi-II to the Kotera and Fidaner methods. We use 16 motives, mostly identical to the ones from the behavioral experimentation: 10 of which depict natural images and 6 represent Ishihara plates.

ViSDEM is implemented with PsychoPy2 [48], and we use multiple Python libraries for the statistical data analysis, namely the NumPy, the Matplotlib, the SciPy, and the Pandas libraries [41, 49–51]. We conduct the experiments on a PC with Windows 7 calibrated by an Eye-One Match Pro to medium white (D65), a gamma of 2.2 and illuminance of 120 lux. We used D50-like fluorescent lights as surrounding lights that were dimmed to approximately 200 lux resulting in a color temperature of ca. 4230 K for the CVD testing, and dimmed to ca. 30 lux (4411 K) for the evaluation.

4. RESULTS

4.1 Behavioral Results

The accuracy values for color-deficient observers show distinctly different accuracy values as seen in Figure 5a: From Yoshi-II (0.94), over Kotera (0.87) and Fidaner (0.79), to the original (0.70) version. All accuracy values are statistically significantly different from each other (cf. Table 1a). In contrast, the results from the normal-sighted are more homogeneous (cf. Fig. 5b): Starting with Fidaner (0.99) and the original (0.99) to Kotera (0.96) and Yoshi-II (0.96). The χ^2 test reveals that the accuracy values for Kotera and Yoshi-II are statistically significantly lower than the values for Fidaner and the original (cf. Table 1b).

Deutan color-deficient observers have a similar overall ranking as color-deficient observers in total (cf. Fig. 5c): First Yoshi-II (0.95), then Kotera (0.92), Fidaner (0.78) and finally the original (0.74). Only the original and Fidaner are *not* statistically significantly different from each other (cf. Table 1c). However, the ranking changes slightly for protan color-deficient observers (cf. Fig. 5d): The highest accuracy is measured for Yoshi-II (0.90), then Fidaner (0.82), Kotera (0.74) and the original (0.62). All image versions are statistically significantly different from each other (cf. Table 1d).

The accuracy values of color-deficient observer for natural scenes depict only little variances (cf. Fig. 5e): The highest accuracy is measured for Yoshi-II (0.95), then Fidaner (0.92) and the original (0.89), and finally Kotera (0.86). The accuracy value for Yoshi-II is statistically significantly higher than Kotera and the original, whereas the accuracy value for Fidaner is statistically significantly greater than Kotera (cf. Table 1e). The accuracy values for the Ishihara images have a similar pattern as the overall accuracy values but with greater variances (cf. Fig. 5f): Yoshi-II with the highest accuracy value (0.92), then Kotera (0.87), Fidaner (0.57), and the original (0.35). Only Kotera and Yoshi-II are

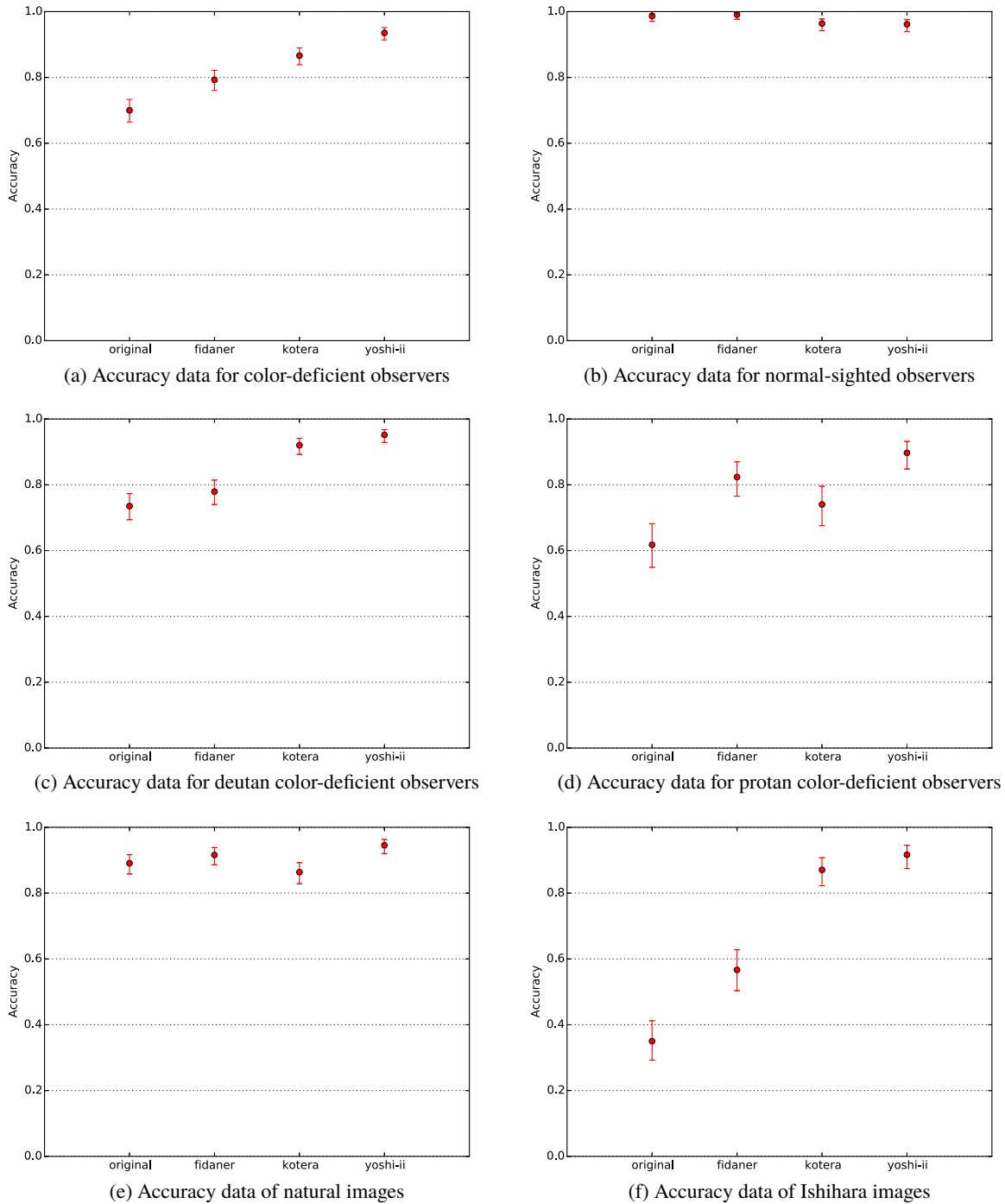


Figure 5. Accuracy data for color-deficient observers versus normal-sighted (top), protan versus deutan color-deficient observers (middle), and natural versus Ishihara images for color-deficient observers (bottom).

not statistically significantly different from each other (cf. Table If).

4.2 Psychometric Results

The ranking from the psychometric experimentation varies slightly. The ranking from the psychometric testing according to preference values, z-scores, can be seen in Figure 6a: It shows the highest preference for Yoshi-II (0.22), Kotera (0.14), original (0.03), and finally Fidaner (−0.38). We can

see a similar preference ranking for deutan color-deficient observers (cf. Fig. 6c): This time starting with Kotera (0.33), then Yoshi-II (0.30), then the original (−0.08), and finally Fidaner (−0.55). The preferences for protan color-deficient observers shifts (cf. Fig. 6d): First the original (0.36), then Fidaner (0.12), Yoshi-II (−0.02), and finally Kotera (−0.46). Normal-sighted observers have a somewhat different ranking (cf. Fig. 6b): The original with the highest

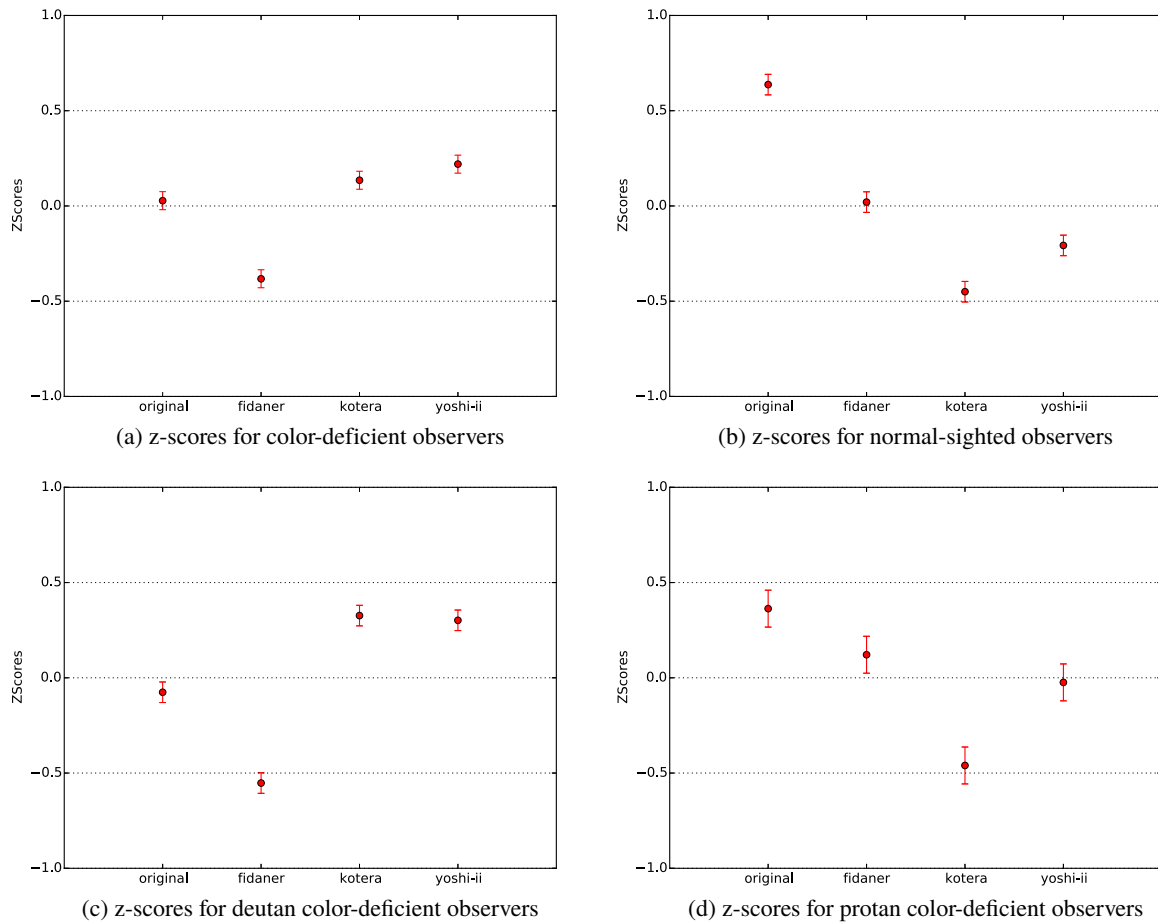


Figure 6. Z-scores of different observer groups for the investigated daltonization method. The z-scores represent the worth value of a daltonization method. A high z-score indicates that observers prefer the daltonization method more often than the others.

z-score value (0.64), followed by Fidaner (0.02), and finishing with Yoshi-II (-0.15) and Kotera (-0.39).

5. DISCUSSION

The visual examples for deutan daltonization in Fig. 3 show how Yoshi-II reintroduces lost information by adapting confusion colors in hue and chroma. The color differences of the wrestling jerseys, for example, become clearly visible, the berries pop up against the foliage in the background, and the number in the Ishihara plate becomes more readable for the color-deficient observer. We can observe an almost identical effect of the protan daltonization in Fig. 4 for the natural images. The protan daltonization of Ishihara plates, however, highlights some downsides of the algorithm: Although the colors of the dots that form the number become better readable than before, the colors of the other dots change as well to a degree that might distract from identifying the correct number.

However, Ishihara images are not representative of most natural images. Natural images have color gradients, and colored areas are usually not separated by a white background. In natural images, like photographs, our proposed method performs very well. Yoshi-II does not have difficulties with images where confusion colors are

directly adjacent to each other, as can be seen with the berry pictures. Likewise, multiscale can overcome the inherent challenges of computations in the gradient domain in cases where confusion colors are not adjacent to each other, like the jerseys of the wrestlers. Multiscale can even help to overcome some difficulties in images like the Ishihara plates.

Furthermore, the background of the Ishihara plates and most neutral and dark areas in the wrestler images are still white or neutral like in the original thanks to neutral data attachment. Without data attachment, the color changes in these regions would have been more drastic. We assume that data attachment contributes to the high preference and good behavioral results because most color-deficient observers perceive neutral colors in the same way as normal-sighted people do [26]. Many of our color-deficient observers noted, for example, that the color changes in skin color of the wrestler images were somewhat distracting and unappealing. These remarks lead us to believe that Yoshi-II could further be improved by widening the data attachment to cover other memory colors as well like skin, grass, sky, etc. In future work, we could, for example, include a data attachment link based on skin color detection (as in [52]). This skin data attachment would keep skin colors as natural as possible leading to a possible boost in the preference z-score. However, adding

complex skin or memory color identification algorithms might make the daltonization methods too complicated or computational expensive. Another option might be to define a lookup table for a set of memory colors. This should be combined with some sort of color constancy or white balancing algorithm to take different illumination situations into account.

One notable side effect of our proposed method is artifacts introduced by the gradient descent of the reintegration process. Color changes are diffused starting from the edges of objects into the center of said objects. Usually, this diffusion works well for small objects, and the reintegration process can change colors quickly like for the berries. If the optimization stops too early, however, halo effects arise as can be seen on the edges of the wrestler image and in the leg area of the left wrestler. We could address these problems by using different functions for the optimization function like total variational or anisotropic approaches that might both reduce halo effects at the edges. Total variation (as in [53]) will minimize the strength of diffusion at the borders at the same time as colors might still be diffused across the edges. Anisotropic diffusion (as in [54, 55]) will not only reduce diffusion *at* the edges but will only diffuse the colors *along* the edges never *across*.

The results from the behavioral and psychometric evaluation reveal a clear ranking for color-deficient observers (cf. Figs. 5a and 6a). We can rank the daltonization methods from highest to lowest accuracy as following:

- (i) From Yoshi-II to
- (ii) Kotera,
- (iii) Fidaner, and
- (iv) the original.

This ranking of the two preexisting methods, Kotera and Fidaner, and the original versions confirms the ranking we observed in [23]. Moreover, we can rank the daltonization methods according to preference from

- (i) Yoshi-II to
- (ii) Kotera,
- (iii) the original, and
- (iv) Fidaner.

The results show that our proposed method, Yoshi-II, not only increases behavioral response for color-deficient observers compared to both the original and the other daltonization methods, but that Yoshi-II is also preferred by the color-deficient observers to the original and any other daltonization method.

The accuracy of Yoshi-II for color-deficient observers in total is not only very high but also lays within the accuracy CI of normal-sighted people (cf. Figs. 5a and 5b). Likewise, the accuracy value of Yoshi-II for normal-sighted observer is still very high, although it decreases slightly as compared to the original. Both observations support the beneficial performance of our proposed method in the light of universal design: Namely, that daltonization improves

behavioral response for the color-deficient at the same time as it restricts deterioration for normal-sighted observers [23].

Yoshi-II performs also best for different types of CVDs (cf. Figs. 5c and 5d): Yoshi-II improves the behavioral response as compared to the original for both protan and deutan color-deficient observers. More precisely, its accuracy value ranks among the two highest accuracy values for both protan and deutan color-deficient observers. Fidaner seems to perform better than its average for protan color-deficient observers, whereas Kotera performs better than its average for deutan color-deficient observers. The performance of the Fidaner method is no surprise because it was originally optimized for protanopes [16]. Daltonization using the protan option of the Kotera method often causes yellow tints in areas with neutral or light colors. These tints might distract protan color-deficient observers and might cause the lower accuracy. The results for protan color-deficient observers have to be taken with care, however, because five protan color-deficient observers might not be representative of the total group of protan color-deficient observers even though the results show statistically significant differences. Moreover, the accuracy values for “natural images only” are much closer than the values for “Ishihara images only” (cf. Figs. 5e and 5f) or both values combined as for the results in [23]. At the same time, however, the accuracy value of Yoshi-II is not only higher than the original, but it is also among the two highest values for both image types.

Although Yoshi-II is preferred to the original and any other daltonization method by all color-deficient observers combined (cf. Fig. 6a), the observer groups have different preferences for the investigated methods. On the one hand, Yoshi-II and Kotera are preferred to the original or Fidaner by most deutan color-deficient observers (cf. Fig. 6c). On the other hand, the original ranks highest amongst protan color-deficient observers followed by the Fidaner method (cf. Fig. 6d). The reasons might be similar to the ones discussed above, namely that Fidaner was optimized for protanopes and Kotera creates a yellow tint for some images. However, the results for protan color-deficient observers have to be taken with caution as mentioned before.

Furthermore, there seems to be a correlation between behavioral accuracy and preference for color-deficient observers. The results from the psychometric evaluation, for example, are more or less congruent with the results from the behavioral evaluation with the exception that the preference values of the original version always rank higher than its accuracy values from the behavioral experimentation. The results for normal-sighted observers, in contrast, reveal a much bigger difference between behavioral and psychometric results (cf. Fig. 6b). More precisely, the original is preferred much more than any of the other version, with Fidaner being the second. Many of our observers stated that they judged preference according to how natural they perceived an image. Thus, they might have preferred the original most often because it appears the most natural. The Fidaner ranking as the second is not surprising either since Fidaner changes only the lightness of confusion colors

such that the Fidaner looks very similar to the original. However, naturalness did not seem to be equally important for color-deficient observers.

There are a few fundamental differences between behavioral and psychometric evaluation strategies. The psychometric experiment measures the subjective preference of the observers, whereas the behavioral test assesses the objective reflexive reaction of the observers' HVS [56]. Which one of the evaluation strategies is, therefore, most useful and most trustworthy? A daltonization method is a visual tool for the color-deficient to adapt confusion colors because chromatic edges and contrast are harder to detect, which has consequences on their behavioral responses. In other words, daltonization should have the goal of improving the behavioral response in general, and chromatic edges and contrast in specific. Thus, we argue the behavioral evaluation to be more important than the aesthetic or preferential, i.e., psychometric, assessment of the image.

We chose the psychometric and behavioral evaluation methods because we wanted to get empirical results from human observers with actual CVDs. We could have considered another objective evaluation method that could involve saliency models before and after daltonization. If the average saliency, for example, increased in the simulated image after daltonization compared to before daltonization, then the daltonization algorithm would be performing as it should be. Another strategy could be to measure color differences between adjacent areas in the simulated image before and after daltonization. The advantage would be that we could get a general idea of improvement without human observers. At the same time, however, the results would be influenced by the unknown accuracy of the simulation methods.

6. CONCLUSION

We introduce a daltonization method for color-deficient people focusing on the enhancement of chromatic edges and contrast called Yoshi-II. Our proposed method is the first method that enhances images in the gradient domain, which is well suited for detecting chromatic edges and contrast. We obtain very good visual enhancement by rotating and scaling the error between the gradients of the original and its simulation in the color space into the plane of optimal visibility that is orthogonal to the main direction of lost information and the direction of lightness. Furthermore, we present the possibility of multiscaling to improve chromatic edges and contrast of areas that are not directly adjacent to each other. Last, we introduce an option to maintain naturalness by providing an interface for data attachment linked to memory colors like, for example, neutral colors.

Yoshi-II performs very well for different CVD and image types. Our proposed method causes increased behavioral response in color-deficient observers combined, and protan and deutan observers individually. Behavioral tests also reveal optimal results for both natural and Ishihara images. More precisely, Yoshi-II leads to improved behavioral response among color-deficient observers as compared to the original, on the one hand, and the Kotera and Fidaner dal-

tonization methods, on the other hand. Moreover, Yoshi-II is preferred to the original and any other of the investigated daltonization methods in a psychometric evaluation. It is ranked among the most preferred daltonization methods for both protan and deutan color-deficient observers.

In future work, we will investigate other optimization strategies like total variational and anisotropic solutions to limit halo artifacts. We will also research various data attachment modules to maintain other memory colors like skin colors.

ACKNOWLEDGMENT

The authors would like to thank Dr. Marius Pedersen (NTNU) for his help in setting up the psychometric experiment, and Prof. Phil Green (NTNU) for contributing to the analysis the data from the psychometric evaluation. This research has been funded by the Research Council of Norway through project no. 221073 "HyPerCept—Colour and quality in higher dimensions."

REFERENCES

- J. Neitz and M. Neitz, "The genetics of normal and defective color vision," *Vision* **51**, 633–651 (2010).
- C. Rigden, "The Eye of the Beholder' - Designing for Colour-Blind Users," *Br. Telecommun. Eng.* **17**, 2–6 (1999).
- E. Hansen, *Fargeblindhet*, 1st ed. (Gyldendal Norsk Forlag AS, Oslo, Norway, 2010).
- A. Valberg, *Lys Syn Farge*, 1st ed. (Tapir Forlag, Chichester, England, United Kingdom, 1998).
- D. Flatla and C. Gutwin, "So That's What You See!" Building Understanding with Personalized Simulations of Colour Vision Deficiency," *ASSETS '12: The Proc. 14th Int'l. ACM SIGACCESS Conf. on Computers and Accessibility* (Association for Computing Machinery (ACM), Boulder, Colorado, USA, 2012), pp. 167–174.
- D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.* **195**, 215–243 (1968).
- A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn. Psychol.* **12**, 97–136 (1980).
- J. Simon-Liedtke and I. Farup, "Empirical Disadvantages for Color-Deficient People," *Mid-Term Meeting of the Int'l. Colour Association (AIC 2015)* (International Colour Association (AIC), Tokyo, Japan, 2015), pp. 391–394.
- M. A. Changizi, Q. Zhang, and S. Shimojo, "Bare skin, blood and the evolution of primate colour vision," *Biol. Lett.* **2**, 217–221 (2006).
- I. Bramão, L. Faisca, K. M. Petersson, and A. Reis, "The Contribution of Color to Object Recognition," *Advances in Object Recognition Systems* (InTech, Rijeka, Croatia, 2012), pp. 73–88.
- N. Milić, D. Novaković, and B. Milosavljević, "Enhancement of image content for observers with colour vision deficiencies," *Color Image and Video Enhancement* (Springer, Cham, 2015), pp. 315–343.
- J.-B. Huang, C.-S. Chen, T.-C. Jen, and S.-J. Wang, "Image recolorization for the colorblind," *IEEE Int'l. Conf. on Acoustics, Speech and Signal Processing, 2009 (ICASSP 2009)* (IEEE, Piscataway, NJ, 2009), pp. 1161–1164.
- G. R. Kuhn, M. M. Oliveira, and L. A. Fernandes, "An efficient naturalness-preserving image-recoloring method for dichromats," *IEEE Trans. Vis. Comput. Graphics* **14**, 1747–1754 (2008).
- G. M. Machado and M. M. Oliveira, "Real-time temporal-coherent color contrast enhancement for dichromats," *Comput. Graph. Forum* **29**, 933–942 (2010).
- H. Kotera, "Optimal daltonization by spectral shift for dichromatic vision," *Proc. IS&T/SID CIC20: Twentieth Color and Imaging Conf. (IS&T, Springfield, VA, 2012)*, pp. 302–308.

- ¹⁶ C.-N. Anagnostopoulos, G. Tsekouras, I. Anagnostopoulos, and C. Kalloniatis, "Intelligent modification for the daltonization process of digitized paintings," *Proc. 5th Int'l. Conf. on Computer Vision Systems* (Universität Bielefeld, Bielefeld, Germany, March 2007).
- ¹⁷ A. Alsam and M. S. Drew, "Fast Colour2grey," *Proc. IS&T/SID CIC16: Sixteenth Color Imaging Conf.* (IS&T, Springfield, VA, 2008), pp. 342–346.
- ¹⁸ S. Di Zeno, "A note on the gradient of a multi-image," *Comput. Vis. Graph. Image Process.* **33**, 116–125 (1986).
- ¹⁹ H. Goldstein, *Classical Mechanics*, Addison-Wesley World Student Series (Addison-Wesley, Reading, Mass, 1950), Vol. 1.
- ²⁰ A. Tveito, H. P. Langtangen, B. F. Nielsen, and X. Cai, *Elements of Scientific Computing* (Springer, Berlin Heidelberg, 2010), Vol. 7.
- ²¹ J. Simon-Liedtke and J. Y. Hardeberg, "Task-Based Accessibility Measurement of Daltonization Algorithms for Information Graphics," *12th Congress of the Int'l. Colour Association (AIC 2013)* (Int'l. Colour Assn. (AIC), Newcastle, UK, July 2013), pp. 108–111.
- ²² J. T. Simon-Liedtke, I. Farup, and B. Laeng, "Evaluating color deficiency simulation and daltonization methods through visual search and sample-to-match: SaMSEM and ViSEM," *Proc. SPIE* **9395**, 939–513 (2015).
- ²³ J. T. Simon-Liedtke and I. Farup, "Evaluating color vision deficiency daltonization methods using a behavioral visual-search method," *J. Vis. Commun. Image Represent.* **35**, 236–247 (February 2016).
- ²⁴ L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.* **34**, 273–286 (July 1927).
- ²⁵ P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development* (Imcotek Press, Winchester, MA, USA, 2000).
- ²⁶ H. Brettel, F. Viénot, and J. D. Mollon, "Computerized simulation of color appearance for dichromats," *J. Opt. Soc. Amer. A* **14**, 2647–2655 (1997).
- ²⁷ T. Lindeberg, "Scale-space theory: A basic tool for analyzing structures at different scales," *J. Appl. Stat.* **21**, 225–270 (1994).
- ²⁸ J. T. Simon-Liedtke, D. Flatla, and E. N. Bakken, "Checklist for Daltonization methods: Requirements and characteristics of a good recolouring method," *IS&T Electronic Imaging: Color Imaging XXII: Displaying, Processing, Hardcopy, and Applications Proc.* (IS&T, Springfield, VA, 2017), pp. 21–27.
- ²⁹ R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. (Pearson Education Inc., Upper Saddle River, NJ, USA, 2008).
- ³⁰ F. Viénot, H. Brettel, and J. D. Mollon, "Digital video colourmaps for checking the legibility of displays by dichromats," *Color Res. Appl.* **24**, 243–252 (1999).
- ³¹ M. Lucassen and J. Alferdinck, "Dynamic simulation of color blindness for studying color vision requirements in practice," *CGIV2006: 3rd European Conf. on Colour in Graphics, Imaging, and Vision* (IS&T, Springfield, VA, 2006), pp. 355–358.
- ³² J. T. Simon-Liedtke and I. Farup, "Using a behavioral match-to-sample method to evaluate color vision deficiency simulation methods," *J. Imaging Sci. Technol.* **60** (2016).
- ³³ I. Jolliffe, *Principal Component Analysis* (Springer, Berlin, Heidelberg, 2002).
- ³⁴ J. T. Simon-Liedtke and I. Farup, "Spatial Intensity Channel Replacement Daltonization (SiChARDA)," *Proc. SPIE* **9395**, 939–516 (2015).
- ³⁵ International Electrotechnical Commission (IEC), "IEC 61966-2-1:1999," Technical Report, International Electrotechnical Commission (IEC) (1999).
- ³⁶ S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Trans. Image Process.* **15**, 141–159 (2006).
- ³⁷ B. Jähne, *Digital Image Processing*, 6th ed. (Springer-Verlag-Berlin-Heidelberg, Berlin, Germany, 2005).
- ³⁸ L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D* **60**, 259–268 (Nov. 1992).
- ³⁹ A. H.-D. Cheng and D. T. Cheng, "Heritage and early history of the boundary element method," *Eng. Anal. Bound. Elem.* **29**, 268–302 (2005).
- ⁴⁰ G. Sharma, *Digital Color Imaging Handbook* (CRC Press, Inc., Boca Raton, FL, USA, 2002).
- ⁴¹ SciPy Developers, "SciPy documentation," <http://www.scipy.org/>, 2013, Last checked: 04/23/2015.
- ⁴² L. Hardy, G. Rand, and M. C. Rittler, "HRR polychromatic plates," *J. Opt. Soc. Am.* **44**, 509–521 (1954).
- ⁴³ J. Neitz and J. Bailey, *H.R.R. Pseudochromatic Plates*, 4th ed. (Richmond Products, Inc., Albuquerque, NM, USA, 2002), originally developed by Hardy LeGrand, Getrude Rand, and M. Catherine Rittler.
- ⁴⁴ B. W. Keelan and H. Urabe, "ISO 20462: A psychophysical image quality measurement standard," *Image Qual. Syst. Perform.* **5294**, 181–189 (2003).
- ⁴⁵ E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *J. Am. Stat. Assoc.* **22**, 209–212 (1927).
- ⁴⁶ G. G. Lovås, *Statistikk for Universiteter Og Høgskoler*, 2nd ed. (Universitetsforlaget, Oslo, Norway, 2008).
- ⁴⁷ E. D. Montag, "Empirical formula for creating error bars for the method of paired comparison," *J. Electron. Imaging* **15**, 010–502 (2006).
- ⁴⁸ J. Peirce, "Psychopy documentation," <http://www.psychopy.org/index.html>, 2014, Last checked: 04/23/2015.
- ⁴⁹ NumPy Developers, "NumPy documentation," <http://www.numpy.org/>, 2013, Last checked: 04/23/2015.
- ⁵⁰ Matplotlib Development Team, "Matplotlib documentation," <http://matplotlib.org/>, 2014, Last checked: 04/23/2015.
- ⁵¹ PyData Development Team, "Pandas documentation," <http://pandas.pydata.org/>, 2012, Last checked: 04/23/2015.
- ⁵² J. Kovac, P. Peer, and F. Solina, "Human skin color clustering for face detection," *IEEE* **2** (2003).
- ⁵³ P. Blomgren and T. F. Chan, "Color TV: Total Variation Methods for Restoration of Vector-Valued Images," *Image Process. IEEE Trans.* **7**, 304–309 (1998).
- ⁵⁴ G. Sapiro, "Vector-valued active contours," *Proc. 1996 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 1996), pp. 680–685.
- ⁵⁵ G. Sapiro, "Color snakes," *Comput. Vis. Image Underst.* **68**, 247–253 (1997).
- ⁵⁶ D. Purves and R. B. Lotto, *Why We See What We Do Redux: A Wholly Empirical Theory of Vision* (Sinauer Associates, Inc., Sunderland, Massachusetts, USA, 2011).