



# CVD-MET: an image difference metric designed for analysis of color vision deficiency aids

J. MUÑOZ-POSTIGO, E. M. VALERO,   
M. A. MARTÍNEZ-DOMINGO,  L. GOMEZ-ROBLEDO,   
R. HUERTAS, \* AND J. HERNÁNDEZ-ANDRÉS 

*Department of Optics, University of Granada, 18071 Granada, Spain*

\**rhuertas@ugr.es*

**Abstract:** Color vision deficiency (CVD) has gained in relevance in the last decade, with a surge of proposals for aid systems that aim to improve the color discrimination capabilities of CVD subjects. This paper focuses on the proposal of a new metric called CVD-MET, that can evaluate the efficiency and naturalness of these systems through a set of images using a simulation of the subject's vision. In the simulation, the effect of chromatic adaptation is introduced via CIECAM02, which is relevant for the evaluation of passive aids (color filters). To demonstrate the potential of the CVD-MET, an evaluation of a representative set of passive and active aids is carried out both with conventional image quality metrics and with CVD-MET. The results suggest that the active aids (recoloration algorithms) are in general more efficient and produce more natural images, although the changes that are introduced do not shift the CVD's perception of the scene towards the normal observer's perception.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

## 1. Introduction

Contrary to the popular belief, color vision deficiency (CVD) is quite widespread, since around 8% of the male and 0.1% of the female Caucasian population is affected by this condition [1]. The main cause of CVD is a spectral shift in the cone spectral responsivities. Normal color vision observers have a range of peak spectral responses either for the L or the M cone. In anomalous trichromatic color vision, some observers (protan type) have two M cone pigments (both within the range covered by the normal M cone pigment) while others (deutan type) have two L cone pigments [2]. In the most severe cases (dichromatic color vision), the observer only has two cones, the S cone and either the M cone (protanopes) or the L cone (deutanopes). All CVD subjects present a diminished color discrimination capability in comparison with normal color vision subjects [3,4], and they can face some difficulties to carry out daily tasks, or else be considered as not eligible for certain professions [5]. These problems have pushed forward the research on aid systems for CVD, which are generally classified in two categories: passive (like color filters) or active (like smart glasses or recoloration algorithms [6]).

Passive aids are advertised as enhancing color contrast in general for CVD subjects. Indeed, much of this enhancement is actually due to changes in luminance contrast, which explain why passive aids can, for instance, allow even a dichromat observer to pass the Ishihara test or discriminate between two colors in a given scene that were perceived as equal without the filter. However, there is compelling evidence suggesting that they do not in any way render the CVD subject's vision more similar to normal color vision or globally improve the observer's discrimination capability [7–12]. Some passive aid systems aim to be more personalized and allow the subject to choose among a set of colored filters the one that he or she prefers, or the one that can offer better results according to some CVD screening test [13].

Active aids can require the use of additional devices like cameras and displays, and they work by transforming the color distribution of the image (recoloration) of a given scene to make it

easier for the CVD subject to better discriminate its colors. To date, many different approaches to recoloration have been proposed [14], and among the most recent proposals are Tsekouras et al. [15], and Xu et al. [16] in 2021. Most methods consider only anomalous trichromacy in general, and the severity of the anomaly is not adjustable, i.e., they are not personalized CVD aids. Besides, most algorithms operate using a global color transform for the image, while only a few try to transform only colors that are confused by the subject, and leave the rest unaltered [15,17,18].

The evaluation of the quality or efficiency of a given CVD aid system is by no means uniform among the different proposals, ranging from just a mere visual presentation of the results to long psychophysical experiments which make a consistent comparison with a sufficient number of alternatives untenable [19]. Performing psychophysical experiments is not easy in the case of CVD aid systems, because there are inherent conceptual and practical difficulties to the problem of evaluating how well a CVD aid system works [20].

At this point, we introduce some clarification on the terminology used in this study. The word “simulation” will refer to algorithms aiming to reproduce a given scene as would be perceived by a CVD subject. We will use “recoloration” to refer to active aid algorithms. Note that some authors refer to active aids as “daltonization” algorithms, and other authors use “daltonization” to refer to simulation algorithms.

The question about which of the two main systems (passive or active) is the best option remains to our knowledge still unclear, and it is directly linked to the lack of a consistent method for quality evaluation of CVD aids. The main aim of this paper is to tackle both issues by introducing an image quality metric that has been designed specifically for the purpose of CVD aid systems evaluation. The underlying idea of the metric is based on the notion of what a “perfect” CVD aid could potentially achieve: to render the appearance of the scene as viewed by a CVD observer as similar as possible to the appearance for a normal observer; and, at the same time, not be too disruptive for the CVD subject, i.e., not interfere greatly with the natural appearance of the colors of the objects in the scene. We have considered two main terms in this new metric (CVD-MET): the efficiency and the naturalness (see Section 2.5). In consonance with these ideas, an efficient CVD aid should modify in a higher degree the problematic or confusing colors for the CVD subject, while leaving the colors that do not present difficulties for the subject quite unaltered. Then, the CVD-MET can separate the colors into these two categories and makes it possible to carry out an independent analysis for both.

Since the metric will be evaluating the CVD aid in terms of the subject’s perception, it must be linked to a simulation method, and this offers the advantage of making CVD-MET personalized to some degree: one of the parameters will be the CVD degree of severity and type of the subject, which is related to the threshold used to classify the colors into problematic or not. None of the conventional image quality metrics can provide these advantages.

The second aim of the study is to offer some quantitative data on the comparison between passive and active aids; and for this purpose, we will need to consider the integration of chromatic adaptation into the simulation method, in a similar way to what was presented in [7]. However, long-term adaptation and neural plasticity effects [21,22] are still not considered in this study, because there is currently no way to introduce them into visual simulations.

In this work, we have introduced as well a modification of an existing simulation method, Kotera [23], to render it capable of simulating not only dichromats but also anomalous trichromat subjects, and also of being integrated into the framework of the chromatic adaptation model. This adaptation constitutes an additional contribution of the study, described in Section 2.2.

The remainder of the paper is organized as follows: Section 2 is dedicated to a description of the methods used, including the scenes, simulation algorithm, passive aids (filtered scenes), chromatic adaptation model, recoloration algorithms, and metrics used (including a selected set of four conventional image quality metrics and CVD-MET). Section 3 shows the results obtained

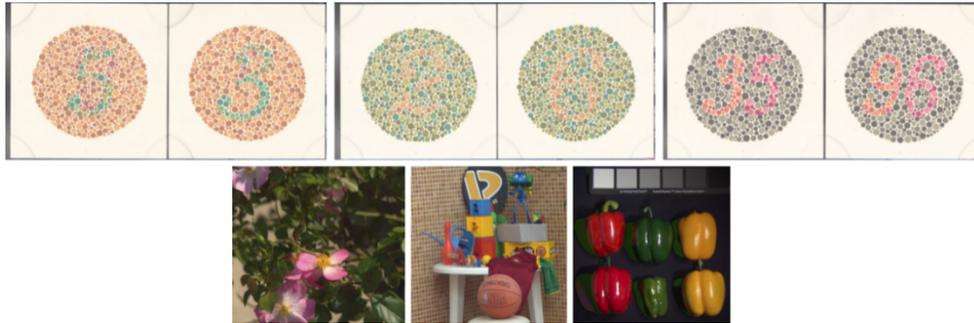
both for the conventional image quality metrics and CVD-MET. Section 4 is dedicated to the discussion of results and comparative analysis. Finally, Section 5 lists the main conclusions of the study.

## 2. Method

### 2.1. Scenes and simulated observers

#### 2.1.1. Hyperspectral scenes

The spectral reflectance information from each pixel in the image is required to obtain a simulation of the scenes as viewed through the colored filters (passive aids). Thus, we have used hyperspectral images from different publicly available databases. The six scenes selected for evaluation of different kinds of CVD aids are shown in Fig. 1. The first three scenes are pairs of plates from the Ishihara test, measured with a push-broom hyperspectral scanner (Pika L, Resonon Ltd., Canada), and available from [7,24]. The fourth and fifth scenes are from the well-known Nascimento, Ferreira and Foster database of rural and urban scenes, and were retrieved from [25,26]. The sixth scene contains three artificial and three natural peppers, and belongs to the CAVE database from Columbia University, retrieved from [27].



**Fig. 1.** Original scenes (images 1-6 from top left to bottom right).

The scenes were interpolated or extrapolated (in the case of Scenes 4 and 5) to the spectral range 400-700 nm, in 5 nm steps. Then, the illuminant D65 was used to obtain the CIE1931 XYZ values of each pixel. The scenes were rendered using the standard XYZ to sRGB transformation [28]. These scenes (see Fig. 1) will be termed as “Original scenes” in the remainder of the paper.

#### 2.1.2. Simulated CVD conditions

As will be described in the next section, we have developed an extension of the Kotera simulation method [23] for anomalous trichromats. In this study, we have simulated three degrees of severity in deutan red-green defects, corresponding to different spectral shifts in M cone responsivity: deuteranope, and two deuteranomalous of different severities. The shift for the deuteranope spans the full distance between the L and M cone responsivities peak wavelengths, for the medium-severe it is eighty per cent of the distance, and for the mild sixty-five per cent of the distance. The cone fundamentals that we have used are those from Smith and Pokorny [29]. The shifts for the anomalous trichromats have been estimated by consulting the data on spectral separation for CVD subjects as described in [30,31], which assume a maximum spectral separation of 27.7 nm between L and M cones, and then translating the separation data to the Smith and Pokorny fundamentals. Therefore, a total of three simulation conditions are analyzed. The way the spectral shifts were applied will be clarified in the next subsection.

## 2.2. CVD simulation method (Kotera2)

The Kotera's simulation method is an extension of the recoloration method which was the main contribution of his study [23,32]. Both recoloration and simulation methods are based on the matrix R decomposition of spectra, which was derived from Wyszecki's hypothesis [33]. Wyszecki postulated that every spectral reflectance can be decomposed into a *fundamental* stimulus that produced the same XYZ values as the original reflectance, and a *metameric black* which produced null XYZ responses. The metameric black corresponds to the portion of the spectral information that is not perceived by the visual system. The decomposition method for obtaining the fundamental spectra was developed by Cohen [34], and can be achieved by using the  $\mathbf{R}_{LMS}$  matrix built from the LMS cone spectral responsivities [23], as shown in Eq. (1):

$$\mathbf{R}_{LMS} = \mathbf{A}_{LMS}(\mathbf{A}_{LMS}^T \mathbf{A}_{LMS})^{-1} \mathbf{A}_{LMS}^T \quad (1)$$

where  $\mathbf{A}_{LMS}$  is a matrix that contains the LMS responsivities. The fundamental spectrum can be obtained by multiplying  $\mathbf{R}_{LMS}$  by the spectral reflectance or color signal. Alternatively, the fundamental spectrum can also be obtained from the tristimulus values through spectral estimation [32]. In this work, the fundamental spectra are obtained from the LMS values by multiplying the  $\mathbf{XYZ}$  matrix of dimensions  $[3 \times N]$ , where  $N$  is the number of pixels in the scene, by the matrix  $\mathbf{P}$ , as shown in Eq. (2):

$$\mathbf{C}^* = \mathbf{P}_{LMS}^{inv} \mathbf{XYZ} = \mathbf{A}_{LMS}(\mathbf{A}_{LMS}^T \mathbf{A}_{LMS})^{-1} \mathbf{LMS} \quad (2)$$

The  $\mathbf{LMS}$  matrix can be obtained from the color signals directly or else from the CIE1931 XYZ tristimulus values, using the standard transformation [28].

Kotera uses a modified  $\mathbf{A}_{LMS}$  matrix for the simulation of a dichromat subject's vision, just incorporating two cone responsivities instead of three. Then, the fundamental spectrum is also obtained for the dichromat ( $\mathbf{C}^*_{DIC}$ ). Kotera's method basic idea is straightforward from this point on: to simulate the dichromat's vision the LMS values are obtained from the  $\mathbf{C}^*_{DIC}$  fundamental spectra by multiplying by the  $\mathbf{A}_{LMS}$  matrix, and following that step, a transformation into the IPT color space [35], applying the null hypothesis for the red-green channel (P) in this space, and transforming back to LMS space. This will produce the dichromat's LMS responses, which can then be remapped into XYZ and sRGB for displaying the simulated scene. A workflow of Kotera's original simulation method is shown in Fig. 2.

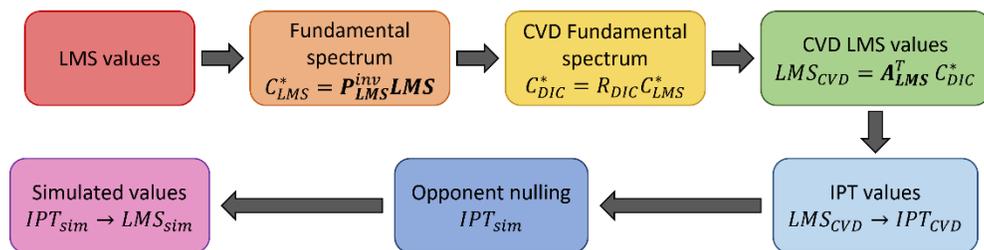


Fig. 2. Workflow of Kotera's simulation method.

We have modified Kotera's original proposal in two ways, to allow us to use the basic structure of Kotera's method for simulations of anomalous trichromats' vision. The first modification is to use three cone responsivities instead of two in the  $\mathbf{A}_{LMS}$  matrix for the CVD observer. In this new  $\mathbf{A}'_{LMS}$  matrix, one of the cone responsivities was shifted from its normal position. The shifted cone responsivity was calculated by displacing the cone pigment responsivity as described in the Appendix of [36]. The amount of the shift was such that, after the corresponding corrections, resulted in the desired shift of the M cone fundamental (65%, 80% or the full span of the L and

M cones spectral separation, see Section 2.1). We have determined the cone pigments shifts by using a Look-Up-Table that related cone pigment shifts with cone fundamental shifts.

The second modification concerns the null hypothesis applied for the red-green (P) channel in the IPT color space. For simulation of an anomalous trichromat, the P channel is multiplied by a factor  $1-\alpha$ , where  $\alpha$  is defined as the ratio between the spectral shift in the cone fundamentals obtained via shifts in the cone pigments,  $\Delta\lambda$ , and the peak distance between the L and M cones of the normal observer. This factor ensures that the channel is only completely nulled if the simulation corresponds to the dichromat case ( $\alpha=1$ ). The normal observer case corresponds to  $\alpha=0$ .

In Fig. 3, we show the simulation results with the Kotera2 method for scene 4 and the three degrees of severity. We can see how the pink areas in the flowers' petals lose chroma as the severity increases.



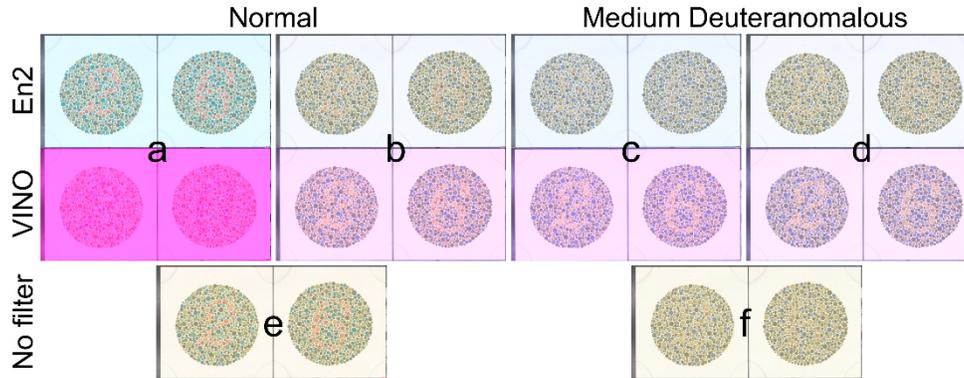
**Fig. 3.** Kotera2 simulations of scene 4 for different CVD conditions: mild deuteranomalous, medium severity deuteranomalous, and deuteranope. The original image is shown on the first row.

### 2.3. Passive aids and CIECAM02 parameters setting

We have selected two colored filters commonly used as passive aids for CVD subjects. One of the filters is from Enchroma (<https://enchroma.com>), indicated for indoor use in deutan observers (Cx1 indoor DT model, En2 from now on). We also have selected the color filter by VINO (<https://www.vino.vi/collections/color-blind-glasses>).

Figure 4(a), shows a sRGB rendering of Scene 2 under En2 and VINO filters, corresponding to the appearance of the scene for a normal subject that had just put on the passive aids (filters). In previous studies [7,8] we have pointed out the importance of considering the effect of chromatic adaptation in simulated scenes with passive aids, and so we have used the well known CIECAM02 Color Appearance Model [37] to render the appearance of the scenes after chromatic adaptation to each filter. The simulated LMS values for each scene and filter were incorporated in CIECAM02, and the computation of the parameters was carried out as described in [7] using white objects present in each scene and two sets of viewing conditions: in a cabin booth with D65 settings for scenes 1-3 and 6, and under outdoor daylight conditions for scenes 4 and 5. In all cases, the  $c$  (impact of background),  $N_c$  (chromatic induction factor) and  $F$  (degree of adaptation) parameters are set to 0.69, 1 and 1, respectively, as proposed default values for average conditions and full adaptation. We have computed the degree of adaptation ( $D$ ) in CIECAM02, using parameters

from Hunt Color Appearance Model [37]. We have obtained the sRGB rendered scenes with the inverse CIECAM02 after chromatic adaptation, as shown in Fig. 4(b) for the normal observer. As can be seen, the effect of the adaptation is very noticeable especially in the background.



**Fig. 4.** Left (a,b,e): Normal observer: (a) Filtered scene 2 with no chromatic adaptation. (b) after chromatic adaptation. (e) with no filter. Right (c,d,f): medium severity deuteranomalous simulation with Kotera2 model: (c) with no chromatic adaptation (d) after chromatic adaptation; (f) with no filter.

To consider chromatic adaptation for CVD subjects, the Kotera2 method has been introduced within the framework of the CIECAM02 model for the filtered images, in the same way as the Lucassen model was used within CIECAM02 model in [7]. The Kotera2 simulation is introduced in the RGB's values of CIECAM02. Figure 4(c) displays the results of the Kotera2 for a medium severe deuteranomalous with no chromatic adaptation, and Fig. 4(d) shows the result of the Kotera2 simulation embedded into CIECAM02. Figure 4(d) aims to show how the scene would appear for this observer after some time wearing the glasses. We can see how the subject would be capable of reading the numbers in the plate with the help of the passive aids especially with VINO filters, in agreement with the results of the psychophysical experiments carried on in [7].

#### 2.4. Active aids (recoloration algorithms)

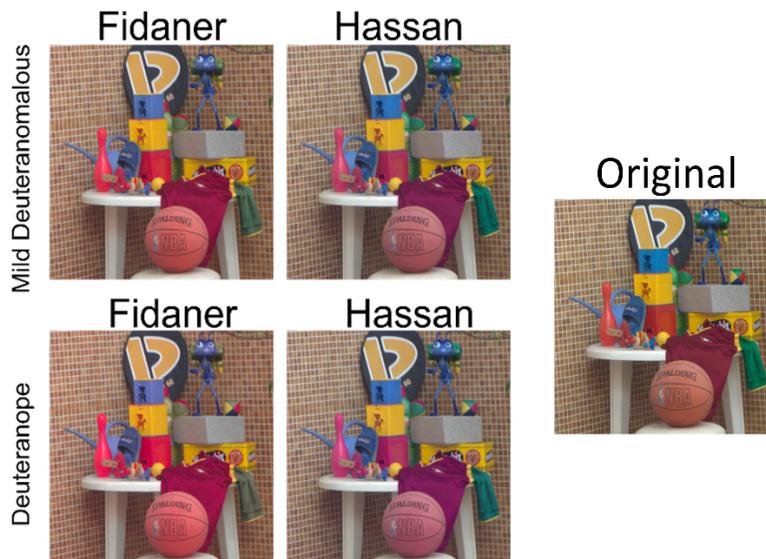
Two recoloration algorithms have been selected with the following criteria: being representative of different design strategies, susceptible of being incorporated into the color appearance framework, and valid for different CVD severities and types. The selected two algorithms are: Fidaner and an adaptation of Hassan and Paramesran [38], which we briefly describe below.

The Fidaner method [39] is based on the idea of obtaining an estimation of the information lost by the CVD subject by subtracting the original scene and the simulated scene in the RGB space, and then re-distributing the lost information among the channels of the recolored image with a specific distribution matrix for each CVD type. Thus, the results will depend on the simulation method applied, and we have used our Kotera2 method in this case. The Fidaner method is then applicable for any severity of red-green anomalies, by changing the simulated image and the distribution matrix according to the type of deficiency (see [39] for details of the distribution matrices used). As mentioned in Section 1, Simon-Liedtke et al. [19] performed a visual-search related evaluation of recoloration methods and in their study the method by Fidaner [40] was the best among those selected.

Hassan and Paramesran's method [38], which we will name as *Hassan* from now on, was originally proposed for dichromats. However, it is easily adaptable for anomalous trichromats, because the first step of the method consists in calculating the difference between the normalized XYZ images corresponding to the original scene and a simulated scene (obtained using Brettel's

method [40] in Hassan's proposal). In our case, we will use the Kotera2 method to obtain the simulated scene, and then the recoloration can be applied for anomalous trichromats as well. The second step in Hassan's method is to find a rotation in the XZ plane to apply in the CIE1931 XYZ space, aiming to transfer the error parameter (obtained as a result of the first step) towards the Z axis. The rotation angle is found as the difference between the angle with the X axis of the XYZ vector of the original pixel and the XYZ vector of the simulated pixel. Once the rotation is applied to the error parameters found in the first step, then the recoloration is built by adding the X rotated error to the X original value, and the sum of rotated X and Z values to the Z original value. Then, the normalization process is reversed, and the XYZ is transformed to sRGB in the final step. It is important to notice that the Y tristimulus value is left unchanged in this method.

Figure 5 (upper row) shows a recolored scene (Scene 5) obtained using Fidaner and Hassan for the mild severity deuteranomalous. Figure 5 (lower row) shows the same scene recolored for the deuteranope case. Hassan's and Fidaner's recoloration look more similar for the mild severity case. We can also appreciate how the recolorations change according to the severity.



**Fig. 5.** Results of active aids for the mild deuteranomalous (Upper row) and deuteranope (lower row) using the Fidaner (left) and Hassan (right) algorithms and scene 5, as seen by a normal observer after recoloration. The original (non-recolored) image is included in the right as reference.

## 2.5. Image quality metrics

The evaluation of performance for CVD aids has been tackled mostly using differently designed psychophysical experiments, but the objective evaluation based on general purpose or specific metrics has been implemented in very few instances [7,8,15]. Even in those instances, the set of metrics selected was not varied or designed specifically for the evaluation of recoloration results or the effect of passive aids on the perceived colors of a given scene.

We have selected a set of three conventional metrics that are able to detect color distortions or alterations in the scene's naturalness, such as those introduced by the CVD aid systems. These metrics were developed within the context of image quality assessment. The purpose of this evaluation is to compare these metrics with CVD-MET in terms of efficiency and naturalness assessment. So, we will use the same image pairs for both the conventional metrics and CVD-MET, but the input images will be sRGB for the conventional metrics, and  $L^*a^*b^*$  values for the

CVD-MET. The two pairs of images being compared are: original vs simulated filtered/recolored (efficiency analysis) and simulated original vs simulated filtered/recolored (naturalness).

With the efficiency analysis, we aim to evaluate how similar to the normal perception of the scene are the recolored or filtered scenes as perceived by the CVD subject. The ideal CVD aid system from the efficiency perspective would make the scene (as viewed by the CVD subject) equal or very close to the normal perception, or at least significantly decrease the difference with the normal subject. Thus, the efficiency term should be as low as possible.

With the naturalness analysis, we intend to evaluate the amount of change introduced by the aid in the CVD perception of the scene. Very intense or brusque changes would likely appear as unnatural for the CVD subject. From the naturalness perspective, the ideal CVD aid system would just enhance contrast in problematic areas and leave non confusing colors unaltered, and so the naturalness term should not be zero (as this would mean that no changes are introduced in the CVD perception of the scene by the aid), but it should not either be very high.

The conventional metrics are not able to directly evaluate the performance or the usefulness of the aid systems. They can just quantify the differences between the pair of images that are being compared. The CVD-MET assesses if the aid system can selectively modify the colors that are confused by the CVD observer, leaving those that are not confused mainly unaltered. The CVD-MET will provide a ranking of the CVD aids based on this perspective and using objective criteria.

### 2.5.1. Conventional image metrics sensitive to color distortions

We have selected four image quality metrics for this evaluation. The selection has been done based on the following two criteria: being sensitive to color distortions in the image as established by the original reference in which the metric was introduced and being representative of different design strategies for quantification of image differences. Three of the metrics are full reference (i.e., they need a reference image apart from the image that is being evaluated), and one is non-reference (i.e., it does not characterize differences between two images, but provides an estimate of the global quality of a given image). The selected metrics are: visual saliency-based index (VSI) [41], multi-scale color image difference metric (MS-iCID) [42], visual information fidelity (VIF) [43], and natural image quality evaluator (NIQE) [44].

VSI includes three components, for quantifying changes in the saliency map obtained with the SDSP model (saliency detection by combining simple priors [41]), a gradient module difference term for detecting changes in contrast, and a chromaticity term for detecting changes in chroma and hue. It preserves symmetry (i.e., the metric is commutative) and the range of values is between 0 and 1, with 1 corresponding to identical images. The VSI metric is computed as shown in Eq. (3):

$$VSI = \frac{\sum_{\Omega} S_{VS}(x) \cdot [S_G(x)]^{\alpha} \cdot [S_C(x)]^{\beta} \cdot VS_m(x)}{\sum_{\Omega} VS_m(x)} \quad (3)$$

where  $S_{VS}$  is the saliency term (modulated by the  $VS_m(x)$  factor, which is the maximum of the two saliency maps at pixel  $x$ ),  $S_G$  is the gradient modulus term, and  $S_C$  is the chrominance term. The weights  $\alpha$  and  $\beta$  are set to 0.41 and 0.02, as recommended in [41]. Given that the optimal value is 1, for our results section we have computed the complement of this metric (1-VSI) and called it cVSI. This will facilitate the comparison of results with the other metrics.

The MS-iCID metric is a multi-scale version of the color image difference (CID) metric [45]. It considers differences in lightness, chroma and hue (including also contrast and structure terms for the first two factors). It is calculated as shown in Eq. (4):

$$MS - iCID = 1 - L_L^5 L_C^1 L_H^1 \prod_{i=1}^n [C_L^i (S_L^i)^3 C_C^i S_C^i]^{\alpha_i} \quad (4)$$

where  $n$  is the number of spatial scales used (5 by default),  $L_L$ ,  $C_L$  and  $S_L$  are the terms related to variations in lightness,  $L_C$ ,  $C_C$  and  $S_C$  the terms related to variations in chroma, and  $L_H$  is related to variations in hue. The  $\alpha_i$  parameters have been set as recommended in [45]. All the terms are computed from the image difference feature maps, after spatially filtering the images and transforming them into the LAB2000HL color space [46]. The MS-iCID metric also preserves symmetry, and its range is from 0 to 1, 0 being the value obtained for two identical images.

The VIF metric [43] quantifies the amount of information that two images have in common. The information is modelled as extracted by a simplified model of the human visual system. The metric computes a decomposition of the image using wavelet-based transforms based on Gaussian Scale Mixtures, and from this decomposition it estimates the amount of information that can be extracted from each image. The metric value is then obtained as a ratio between the amount of information extracted from the test image and its counterpart for the reference image. This metric does not contain explicit chromaticity terms as the other two, but it is proven sensitive to color changes as well as local contrast changes, which makes it a suitable candidate for the evaluation that we are looking for in this study. The range of values is from 0 to above 1 (upper limit not specified), and two identical images will yield a value of 1. Higher than unity values indicate distortions that can be conducive to enhancement, like a contrast increase. Values below unity indicate distortions that cause some loss of information from the reference image. For this metric, we have also computed the complement  $cVIF = (1-VIF)$  in the results section, so that two identical images yield a value of  $cVIF$  equal to zero. This metric does not preserve symmetry, because it is defined as a ratio, as shown in Eq. (5):

$$VIF = \frac{\sum_{j \in \text{subbands}} I\left(\bar{C}^{Nj}; \bar{F}^{Nj} | S^{Nj}\right)}{\sum_{j \in \text{subbands}} I\left(\bar{C}^{Nj}; E^{Nj} | S^{Nj}\right)} \quad (5)$$

The NIQE metric [44] is the only one we have included that is no-reference. It is also distortion-unaware and opinion-unaware, and so it does not require previous knowledge about the kind of distortions present or reflects data from subjective evaluations. The metric is based on obtaining a multi-variant gaussian model of the scene from first order statistics features, and then comparing this model with the one obtained from a set of natural (undistorted) images. The basic idea behind this design is that if an image is undistorted (natural), then the statistical features will follow a Gaussian distribution, while this will not be the case for distorted or unnatural images. Specifically, it is calculated as shown in Eq. (6):

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)^T \left( \frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2)} \quad (6)$$

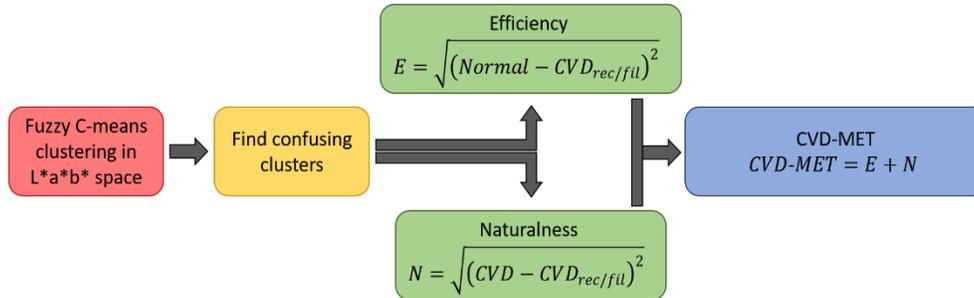
where  $v_1$  and  $\Sigma_1$  are the mean vectors and covariance matrices of the set of natural images, and  $v_2$  and  $\Sigma_2$  the analogue quantities for the image that is evaluated. As it can be deduced from Eq. (6), the NIQE value for a perfectly natural image will be close to zero, and the higher the value of this metric, the less “natural” the image will be. Since we have two images, we have computed the difference between the first and second images’ NIQE values. For instance, for the efficiency comparison, we computed  $NIQE(\text{original}) - NIQE(\text{simul.}_\text{filter}/\text{recolored})$ . Then, a value of 0 would indicate that the naturalness of the recoloration is the same as in the original image, and a negative result would indicate a loss of naturalness after the CVD aid system has been applied.

### 2.5.2. Proposed metric for CVD aids evaluation (CVD-MET)

The CVD-MET metric that is proposed in this paper is composed by the sum of two terms: efficiency and naturalness. Both are based on the use of a CVD simulation model to evaluate the

similarity of a pair of images. As explained at the beginning of Section 2.5, for the efficiency term we will compare the original with the simulated filtered/recolored image, and for the naturalness we will compare the simulated original with the simulated filtered/recolored images. In Section 3, for a more detailed analysis, both naturalness and efficiency are studied independently, omitting the final value of CVD-MET.

In Fig. 6, we present the main steps of the CVD-MET computation.



**Fig. 6.** Workflow of the CVD-MET main computation steps.

The simulation algorithm used will likely influence the results of CVD-MET. The characterization of the behavior of the metric for different simulation algorithms is, however, out of the scope of this study.

One of the main features of the CVD-MET is that the comparison will be carried out in the CIELAB space, and this will allow the results to be interpreted in terms of color differences directly. The other main feature is that the metric is computed separately for colors that are confused by the CVD, and the rest of the colors. A confused color pair is formed by two colors that are perceived as different by a normal observer, and close to equal by the CVD observer. Thus, two thresholds are needed in the CIELAB color space to define what is different for the normal subject and will be considered as close to equal for the CVD subject. The threshold for normal color vision is set to 1.0 CIELAB units, and three thresholds will be considered for the CVD observers depending on their severity: 3 (mild), 4.5 (medium) and 6 (dichromat). The thresholds are just an estimation and could be considered as input parameters for the CVD-MET function, so it can be adapted to different kind of scenes or observers. Obviously, the percentage of total colors that are confused depends on the thresholds. We have chosen the threshold values that allowed around 20% of confused colors in average for all the images. The threshold values could be adapted for a particular individual by measuring the discrimination ellipses in CIELAB for different centers, or else using a custom simulation if the cone responsivities positions are known or can be estimated.

Determining which pair of colors are confused or not is one main advantage of CVD-MET, but it can be quite computationally expensive, especially for large images. We have then introduced a pre-processing color quantization step using Fuzzy C Means clustering in the CIELAB color space (similarly to the Tsekouras recoloration method for dichromats [15], but they clustered in the RGB color space). The optimal number of clusters for each image is determined by using the elbow method [47], that locates the initial point of asymptotic behavior in the curve of clustering quality metric vs number of clusters. The optimal number of clusters ranges from 20 to 35 for the six images analyzed in this study.

### 3. Results

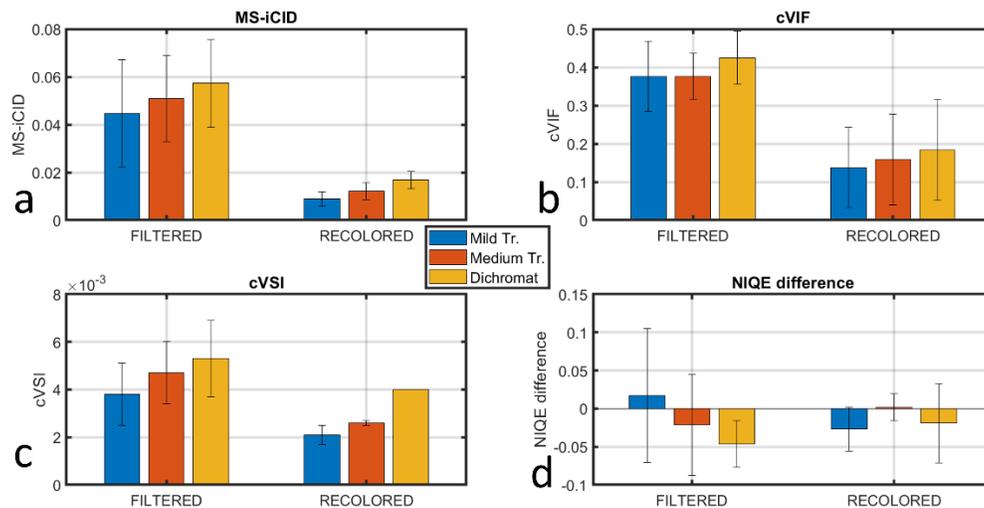
In this section, we will present the results for the efficiency and naturalness terms, both for the conventional metrics and the CVD-MET as defined in Section 2.5. We have analyzed three main

aspects: the influence of the deutan severity, the differences found among algorithms or filters, and the influence of the image selected.

### 3.1. Influence of the CVD condition

To analyze the influence of the CVD severity, we have obtained the average across filters/recoloration algorithms and across images, for each metric.

Figure 7 shows results for the efficiency term for the conventional metrics, which corresponds to the comparison between original and simulated filtered/recolored images. We notice a very high standard deviation in all cases, which is not very surprising because we are averaging over different algorithms and images to obtain the mean results.



**Fig. 7.** Efficiency term metrics results averaged across images and filters/algorithms.

As expected, the trend in Figs. 7(a), 7(c), and for the recolored images in 7(b), is towards less efficiency, i.e., higher values of the metric for the more severe CVD conditions. Thus, the higher the severity of the deutan subject, the harder it is for an aid system to work efficiently. Nevertheless, we must also consider that the loss of chromatic discrimination is also increasing with the CVD condition severity, and this means that the metric value before the aid is applied is also higher for dichromats than for anomalous trichromats.

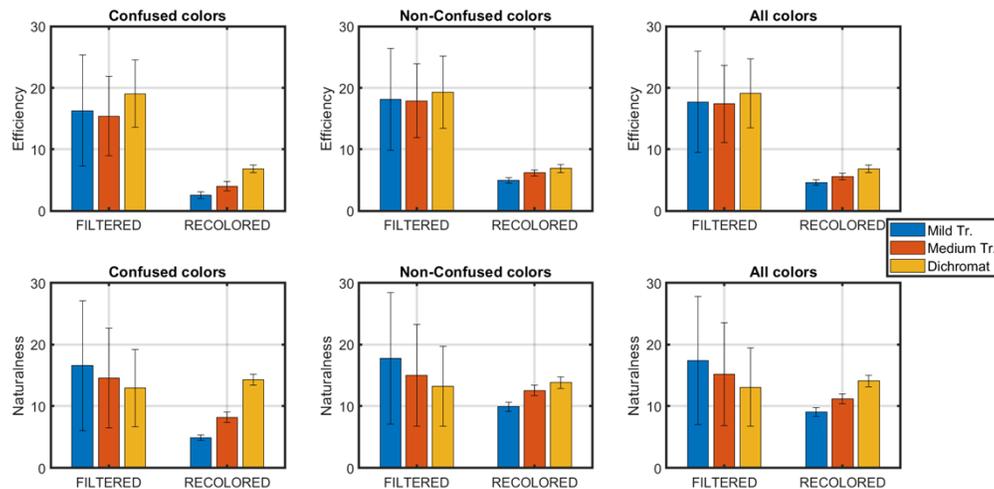
For the filtered images and cVIF metric, Fig. 7(b), there is no clear trend, and the same can be said about the NIQE results, Fig. 7(d). The cVIF metric results show that, from the perspective of information content, the passive aids are not selective according to CVD condition, which can be explained because they do not selectively modify certain areas of the scene differently for different CVD conditions. The recoloration algorithms follow this strategy of selective modification, and so they introduce stronger changes for the higher severity conditions, and this is reflected in slight changes in the cVIF metric's result for the recolored images.

The NIQE results, shown in Fig. 7(d), are more complicated to interpret because the values shown correspond to the difference between the NIQE value of the original image and the NIQE value of the simulated recolored/filtered image. Since lower NIQE values correspond to more natural images, a positive NIQE difference value indicates that the naturalness of the simulated modified image (filtered/recolored) is higher than the original image's; that happens in the mild condition for the passive aids, while the opposite trend is found for the other conditions and for the active aids. The negative sign obtained for most conditions shows that altered images are less natural than the original scenes. It can be concluded that the aids cause some alterations

in the naturalness of the images according to NIQE values, but these changes are not strongly dependent on the severity of the CVD condition.

Regarding the naturalness term, i.e., the comparison between the simulated original and the simulated recolored/filtered images, the trends for the recolored images are very similar to those shown in Fig. 7. These results are not shown due to limitations of space. This suggests that the strongest changes introduced by the recoloration algorithms for dichromats are also altering the appearance of the scene to a higher degree for these observers.

Regarding the CVD-MET results, the data are composed of three parts: confused colors, non-confused colors and global result (considering the full set of colors). In Fig. 8, the results for the CVD condition factor are shown for the efficiency term (upper row) and naturalness term (lower row).



**Fig. 8.** CVD-MET results for the CVD severity factor, obtained by averaging the metric's results over scenes and CVD aids, separately for active (recolored) and passive (filters) aids. Upper row: efficiency term. Lower row: Naturalness term. Left: confused colors. Middle: non-confused colors. Right: all colors.

The influence of the CVD condition is clearly higher for the recolored images, with worse efficiency and naturalness results for the dichromats than for the anomalous trichromats, and the results are ranked by severity. Note that the starting point for deuteranopes is further from the normal image than for deuteranomalous, see Fig. 3. This is not the case for the filtered images and the efficiency term, in which the average results for the anomalous trichromats do not present a clear trend with respect to the dichromats. This is an expected result, because the recoloration algorithms produce a different result for each CVD condition, while the filtered original images do not change. In the naturalness term for the filtered images, there is an opposite trend (CVD-MET values are lower for higher severity), which indicates that adding the filter produces higher differences in the appearance of the scene (as seen by the CVD subject) for the mild than for the severe conditions. This can be explained as consequence of the loss in color discrimination for the more severe CVD conditions, which partially compensates the global color change introduced by the filters.

The average naturalness term values tend to be higher than the efficiency values for the active aids, but the opposite trend is found for the passive aids. This suggests that using CVD aids inevitably comes at the cost of sacrificing the appearance of the scene, and thus it will affect naturalness, but in a higher degree for the active than for the passive aids. The fact that the passive aids introduce a change that goes in a fixed direction in the color space for the whole scene

makes them less efficient but to a certain extent contributes more towards preserving naturalness, especially if the chromatic change is not a huge one. This is the case for the En2 filter, but not so much for the VINO filter, as seen in Fig. 4.

Regarding the division into confused and non-confused colors, the general trends commented above for each term are shared by both groups, but the average efficiency and naturalness values are higher for the non-confused colors (meaning worse results in general for the non-confused colors). This happens also for the recolored images, which suggests that in average the recoloration methods are failing to successfully differentiate between confused and non-confused colors, although the change is less noticeable for the deuteranopes. This is not very surprising because not many of the available recoloration methods are based in a design strategy that prioritizes this distinction, and neither Fidaner nor Hassan algorithms are strictly differentiating between confusing and not confusing colors.

The results of the efficiency term in CVD-MET for this factor agree with the cVIF metric results (see Fig. 7), although the relative efficiency of the active aids in comparison to the passive aids seems to be higher, meaning that they produce lower metric values.

### 3.2. Influence of the recoloration algorithm/filter

To study this factor, we will average the results across different images and different CVD conditions. The results are shown in Table 1 for the filtered images and recolored images.

**Table 1. Metrics for filtered and recolored images. Average data (and standard deviation) for different filters/algorithms. Green and blue for lowest and highest values respectively.**

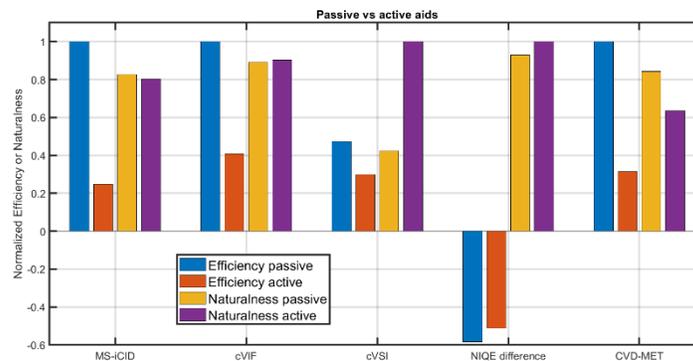
Efficiency	En2	VIN	Fidaner	Hassan
Ms-iCID	0.0313(0.0339)	0.0706(0.0722)	0.0160(0.0169)	0.0092(0.0102)
CVIF	0.3182(0.0497)	0.4662(0.1651)	0.2785(0.1294)	0.0409(0.0548)
CVSI	0.0032(0.0022)	0.0060(0.0019)	0.0031(0.0026)	0.0027(0.0023)
NIQE difference	-0.0783(0.1074)	0.0446(0.1239)	-0.0476(0.1101)	0.0181(0.0342)
CVD-MET conf	10.0904(3.9226)	23.8317(10.6487)	5.2143(3.1246)	3.9575(3.0580)
CVD-MET no-conf	11.7277(2.9577)	25.1695(8.4239)	6.5723(2.3132)	5.4546(2.7152)
CVD-MET all	11.3425(2.0822)	24.7721(8.9941)	6.2206(1.7993)	5.1430(2.3945)
Naturalness	En2	VIN	Fidaner	Hassan
Ms-iCID	0.0142(0.0145)	0.0700(0.0743)	0.0450(0.0462)	0.0368(0.0390)
CVIF	0.2145(0.0613)	0.4841(0.1705)	0.4285(0.1550)	0.2798(0.0842)
CVSI	0.0006(0.0007)	0.0077(0.0039)	0.0103(0.0103)	0.0092(0.0090)
NIQE difference	-0.0346(0.0677)	0.0883(0.1281)	-0.0040(0.1117)	0.0617(0.0859)
CVD-MET conf	6.4457(2.2903)	22.7177(11.8296)	10.1310(7.8743)	8.6198(6.1769)
CVD-MET no-conf	6.8848(0.9950)	23.7909(10.3449)	12.9592(5.8679)	11.2563(5.2393)
CVD-MET all	6.8464(1.2035)	23.5842(10.6795)	12.2714(4.9089)	10.6402(4.1472)

We will first describe the results for the passive aids, which are mostly consistent across metrics and show that the filter that offers the best results for the efficiency and naturalness terms is En2. The results of the naturalness and efficiency agree, and for the conventional metrics tend to be similar for the efficiency and naturalness terms (although lower for naturalness in the En2 filter case). For CVD-MET results, the naturalness values tend to be lower for the passive aids but higher for the active aids, as commented in the previous section.

Globally, the best performing active aid is Hassan, while Fidaner is less efficient and offers less-natural images, according to most metrics. Higher values of CVD-MET for the non-confused colors can also be observed here.

Comparing the values for passive and active aids as shown in Table 1, from the efficiency point of view, the active aids are superior according to all metrics. Regarding the information content (assessed by the cVIF metric), there is also a higher similarity between the original and recolored images, as perceived by the CVD observer, than between the original and filtered images. From the naturalness point of view the En2 filter is best according to all metrics except NIQE difference. The VINO filter offers the worst results in all metrics. This can be explained if we consider that the En2 filter introduces only a slight chromatic change in the scene (partially compensated by chromatic adaptation), while the VINO filter tends to increase the saturation to a higher extent. In terms of preserving the naturalness of the original scene, the VINO filter and the Hassan algorithm tend to perform worse. The results suggest that filters that introduce slight changes do so at the expense of effectiveness, while active aids that are more effective produce a less natural appearance in the scenes.

Figure 9 shows the normalized average naturalness and efficiency values for active and passive aids. Each group of bars has been normalized independently to their maximum value for an easier comparison. All metrics show higher efficiency (lower metric values) for the active than for the passive aids. The naturalness results are closer between active and passive aids (save for the cVSI, as explained before), and the trends differ between metrics, showing that there is not a clear agreement in average on which type of aid provides more natural images. As we discussed earlier, this is influenced by the marked difference in the behavior of the two passive aids selected.



**Fig. 9.** Comparison of average normalized metrics values for passive vs active aids.

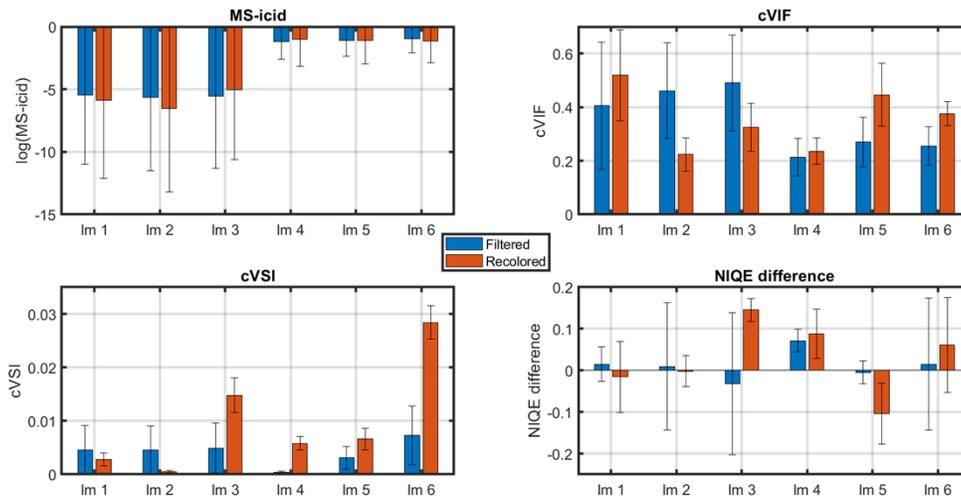
### 3.3. Influence of the image

#### 3.3.1. Conventional metrics

Among the selected images, the Ishihara subgroup has similar characteristics, other two are natural scenes, and the remaining one is an artificial scene in a cabin booth with both natural and man-made objects (see Fig. 1). In this subsection, the influence of the type of images is analyzed. For this purpose, we have computed the average of the metrics across algorithms/filters and across CVD conditions.

Figure 10 shows the averaged values for the naturalness term and the four conventional metrics for each image and separately for the recolored and filtered images. The MS-iCID values for the Ishihara images were very low, so we have used logarithmic scale to represent them.

The MS-iCID metric shows similar values for the three Ishihara images, but the MS-iCID values are much lower than for the natural scenes and the pepper scene. The results for the MS-iCID metric can be conditioned by the fact that the Ishihara images have very large uniform areas and not very clearly defined edges, which would make the contrast terms in MS-iCID have low values.



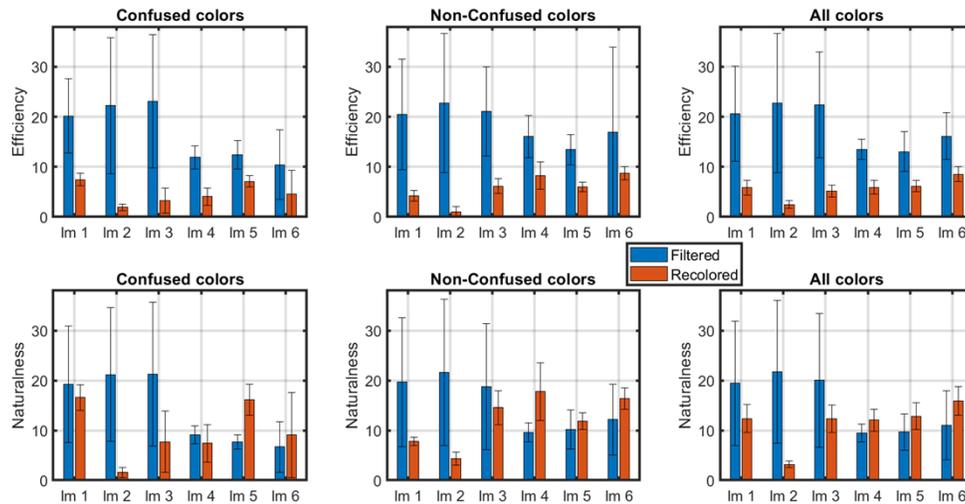
**Fig. 10.** Average of the naturalness term across algorithms/filters and across CVD conditions, for each of the six images used. Conventional metrics.

The cVIF metric values for the filtered images tend to be different and higher for Images 1 to 3 (Ishihara diagnostic pages) while the NIQE values show a variation in sign across the first three images and higher absolute values for Images 3 to 5.

The recolored images tend to present higher values of cVSI (except Scene 2), meaning that they introduce changes that influence saliency to a greater extent, as expected. Also, the fact that we use the simulated filtered images under chromatic adaptation for evaluating the naturalness term tends to decrease the differences in color.

For the efficiency term, the general trends outlined in the previous paragraphs are also present.

Regarding the CVD-MET results, Fig. 11 shows the efficiency (upper row) and naturalness (lower row) values obtained in average for each image with the active and passive aids systems studied.



**Fig. 11.** CVD-MET results for the image factor. Upper row: efficiency term. Lower row: naturalness term. Left: confused colors. Middle: non-confused colors. Right: all colors.

We find different trends for the dependence with the image for filtered and recolored images, and for confused and non-confused colors. The passive aids for the Ishihara plates and the confused colors show less efficiency and naturalness than for the other images, while the best results on average for the active aids were for Image 2 (an Ishihara plate, see Fig. 1). On average, the efficiency for the recolored images is 4.20 for the Ishihara plates, and 5.25 for the natural scenes, which are quite close values. For the filtered images and the confused colors, the averages of the efficiency are 21.83 and 11.58 for the Ishihara plates and the natural images respectively. Thus, for the confused colors, the relative distance in efficiency between passive and active aids is higher for the Ishihara plates than for the natural scenes. In this sense, the active aids would be more clearly outperforming the passive aids for the Ishihara plates.

For the non-confused colors, the efficiency term for the passive aids tends to be higher than the naturalness term for scenes 4-6. The passive aids are as well less efficient for the natural images than the active aids.

This last trend is also supported by the naturalness results, both for confused and non-confused colors. The active aids show less variation in the naturalness across images.

This dependence of the results with the image was expected because the color gamut and the spatial structure depend on the scene contents, and one or both factors will influence any image quality metric.

#### 4. Discussion

In this section, we will present some additional insight derived from the interpretation of the results presented in Section 3.

Among the four conventional metrics analyzed, CVD-MET has in general the higher degree of correlation with the cVIF metric, especially for the deutan severity and image factors. In many instances there is also correlation with MS-iCID. The cVIF metric contains a simplified model of the visual system which is used to estimate the amount of common information between the two images compared. It has proven sensitive to color variations, so it is maybe expected that it would correlate with a metric based on color differences. The MS-iCID metric includes terms specifically related to lightness, chroma and hue differences, so it would be expected a good correlation with the CVD-MET. However, the fact that MS-iCID works with different spatial scales has proven to be determinant in causing a very marked dependence with the spatial characteristics of the images being analyzed (see Fig. 10). This dependence is not so emphasized for the CVD-MET, which due to the clustering pre-processing step are independent of the spatial content of the scene, up to a certain degree.

The fact that the cVSI metric uses a saliency model as a key step has proven beneficial to pinpoint the active aids capability of selectively enhancing certain areas of the image, but considering visual attention is not an optimal design strategy for the evaluation of CVD aids. This metric is mostly aiming to detect visual artifacts in images, and thus it might lead to wrong conclusions at first sight.

The results of NIQE difference are a bit more erratic and difficult to interpret. NIQE is a non-reference metric and is also conditioned by the set of images used to train the algorithm that produces the naturalness index. It is interesting that the NIQE difference efficiency global average value for the active and passive aids is negative, suggesting that the simulated recolored/filtered scenes are less natural than the original scenes. But at the same time, the NIQE naturalness result for all CVD aids is positive, which would be contrary to expectations because it suggests that the recolored/filtered scenes are perceived as more natural by the CVD subject. However, we must consider that the naturalness comparison is performed using only simulated images, and so the color distribution of these images is rather different from that of the original scenes and the scenes used to train NIQE.

Using a set of different metrics is an asset of the evaluation presented in this study, since the differences in ranking or general trends between the metrics can many times be explained in terms of how the metrics are designed or what specific features are they sensitive to. CVD-MET has proven to be quite consistent with conventional metrics in the analysis of the three factors shown in Section 3, and in this sense using additional metrics is a necessary step in the validation of the new metric's results.

## 5. Conclusions

We have introduced a new metric, CVD-MET, designed for CVD aid systems evaluation and performed some experiments to demonstrate its capabilities, assets, and limitations for a set of six images. Some representative instances of active (recoloration algorithms) and passive (filters) aid systems have been evaluated. The dependence of CVD-MET on the CVD severity within the deutan type, algorithm/filter and image have been analyzed in comparison to conventional image quality metrics. Since the CVD-MET design uses two different comparisons for the efficiency and naturalness terms, these two comparisons have also been carried out using the conventional metrics.

These three factors (CVD severity, algorithm/filter and image) have proven to affect the metrics, as expected. The efficiency and naturalness are lower for the deuteranope and are ranked by CVD severity for the active aids. Nevertheless, this conclusion must be interpreted carefully because the initial differences between the normal subject's and the dichromat subject's perception of any scene are higher than for the less severe conditions. Regarding the algorithm/filter factor, the aim of this analysis was not to specifically pinpoint particular algorithms or filters (since we have analyzed just a few representative instances) but to derive more general conclusions about the efficiency of active vs passive aids. In this sense, the consensus result is that active aids are more efficient and produce images that would present less color changes to the observer. However, the long-term neural plasticity effects on the subjects that filter usage can produce are not considered in our simulations [21,22] due to lack of appropriate models. Neural plasticity considerations could potentially result in a better acceptance of the passive aids by the CVD subjects over time. The active aids do not require a continuous usage.

CVD-MET has the advantage over conventional metrics of being designed specifically for the evaluation of CVD aid systems. It is based on color differences, and this makes more sense for the purpose than considering spatial variations of the appearance of significant textural artifacts that should not be present in neither passive nor well-designed active aids. It also offers the possibility of evaluating separately the confused and non-confused colors in the image. The classification of colors into these categories can be done for a particular observer, either by measuring discrimination thresholds around different representative color centers, or else using an experiment of color ranking that would allow an indirect estimation of the average threshold for the subject. In this sense, the CVD-MET can be customized, and extended for many different application fields, like web design, videogames customization, design of emergency signs, maps, etc.

The main limitations of the CVD-MET are two: it is based on simulation models, and thus its results are necessarily linked to the particular simulation model used; and it is to a certain degree parametric regarding the classification of colors into confused and non-confused. The personal preference of CVD subjects about a given aid system color transformation have not been incorporated to the metric, because they are highly dependent on the individual experience and out of the scope of the metric's purpose.

For future improvements of the CVD-MET, we will consider an enhanced method for the classification of colors into confused and non-confused, that is not based on a unique threshold value, and the possibility of using more sophisticated simulation models and evaluation of differences in the subject's own color space instead of the CIELAB space. However, the current

lack of development of CVD color spaces does not allow us to introduce this improvement in the design of the CVD metrics, and the fact that it would complicate the inter-individual comparisons needs to be considered as well.

**Funding.** Junta de Andalucía (A-TIC-050-UGR18); Ministerio de Economía y Competitividad (FIS2017- 89258-P); Ministerio de Ciencia, Innovación y Universidades (RTI2018-094738-B-I00).

**Acknowledgments.** The authors thank the reviewer and guest editor of this special issue for their insight, which has contributed greatly to enhance the Kotera2 simulation method.

**Disclosures.** The authors declare no conflicts of interest.

**Data Availability.** Data and code underlying the results presented in this paper may be obtained from the authors upon reasonable request.

## References

1. J. Birch, "Worldwide prevalence of red-green color deficiency," *J. Opt. Soc. Am. A* **29**(3), 313–320 (2012).
2. M. P. Simunovic, "Colour vision deficiency," *Eye* **24**(5), 747–755 (2010).
3. J. M. Linhares, P. D. Pinto, and S. M. Nascimento, "The number of discernible colors perceived by dichromats in natural scenes and the effects of colored lenses," *Vis. Neurosci.* **25**(3), 493–499 (2008).
4. R. C. Pastilha, J. M. M. Linhares, A. E. Gomes, J. L. A. Santos, V. M. N. de Almeida, and S. M. C. Nascimento, "The colors of natural scenes benefit dichromats," *Vision Res.* **158**, 40–48 (2019).
5. B. L. Cole, "The handicap of abnormal colour vision," *Clin. Exp. Optom.* **87**(4-5), 258–275 (2004).
6. A. Popleteev, N. Louveton, and R. McCall, "Colorizer: smart glasses aid for the colorblind," in *Proceedings of the 2015 workshop on Wearable Systems and Applications*, (2015), pp. 7–8.
7. M. A. Martínez-Domingo, L. Gomez-Robledo, E. M. Valero, R. Huertas, J. Hernandez-Andres, S. Ezpeleta, and E. Hita, "Assessment of VINO filters for correcting red-green Color Vision Deficiency," *Opt. Express* **27**(13), 17954–17967 (2019).
8. L. Gomez-Robledo, E. M. Valero, R. Huertas, M. A. Martínez-Domingo, and J. Hernandez-Andres, "Do EnChroma glasses improve color vision for colorblind subjects?" *Opt. Express* **26**(22), 28693–28703 (2018).
9. H. A. Swarbrick, P. Nguyen, T. Nguyen, and P. Pham, "The ChromaGen contact lens system: colour vision test results and subjective responses," *Ophthalmic Physiol. Opt.* **21**(3), 182–196 (2001).
10. R. Mastey, E. J. Patterson, P. Summerfelt, J. Luther, J. Neitz, M. Neitz, and J. Carroll, "Effect of "color-correcting glasses" on chromatic discrimination in subjects with congenital color vision deficiency," *Invest. Ophthalmol. Vis. Sci.* **57**, 192 (2016).
11. E. J. Patterson, "Glasses for the colorblind: their effect on chromatic discrimination in subjects with congenital red-green color vision deficiency," in *International Conference on Computer Vision Systems (ICVS)*, (2017), pp. 18–22.
12. N. Almutairi, J. Kundart, N. Muthuramalingam, J. Hayes, K. Citek, and S. Aljohani, "Assessment of Enchroma Filter for Correcting Color Vision Deficiency," Pacific University (Oregon) (2017).
13. K. Wenzel and Á Urbin, "Improving colour vision," in *Lumen V4 Conference, Budapest*, (MEE Lighting Society, 2014), pp. 427–438.
14. M. Ribeiro and A. J. Gomes, "Recoloring algorithms for colorblind people: A survey," *ACM Comput. Surv.* **52**(4), 1–37 (2020).
15. G. E. Tsekouras, A. Rigos, S. Chatzistamatis, J. Tsimikas, K. Kotis, G. Caridakis, and C. N. Anagnostopoulos, "A Novel Approach to Image Recoloring for Color Vision Deficiency," *Sensors* **21**(8), 2740 (2021).
16. L. Xu, Q. Li, X. Liu, Q. Xu, and M. R. Luo, "Gamut mapping based image enhancement algorithm for color deficiencies," *Biomed. Opt. Express* **12**(11), 6882–6896 (2021).
17. D. R. Flatla, K. Reinecke, C. Gutwin, and K. Z. Gajos, "SPRWeb: Preserving subjective responses to website colour schemes through automatic recolouring," in *Proceedings of the SIGCHI conference on human factors in computing systems*, (2013), pp. 2069–2078.
18. N. Milić, M. Hoffmann, T. Tómacs, D. Novaković, and B. Milosavljević, "A content-dependent naturalness-preserving daltonization method for dichromatic and anomalous trichromatic color vision deficiencies," *J Imaging Sci Technol* **59**(1), 10504-1 (2015).
19. J. T. Simon-Liedtke and I. Farup, "Evaluating color vision deficiency daltonization methods using a behavioral visual-search method," *J. Vis. Commun. Image Representation* **35**, 236–247 (2016).
20. E. M. Valero, R. Huertas, M. A. Martínez-Domingo, L. Gómez-Robledo, J. Hernández-Andrés, J. L. Nieves, and J. Romero, "Is it really possible to compensate for colour blindness with a filter?" *Color. Technol.* **137**(1), 64–67 (2021).
21. J. S. Werner, B. Marsh-Armstrong, and K. Knoblauch, "Adaptive Changes in Color Vision from Long-Term Filter Usage in Anomalous but Not Normal Trichromacy," *Curr. Biol.* **30**(15), 3011–3015.e4 (2020).
22. J. Neitz, J. Carroll, Y. Yamauchi, M. Neitz, and D. R. Williams, "Color perception is mediated by a plastic neural mechanism that is adjustable in adults," *Neuron* **35**(4), 783–792 (2002).
23. H. Kotera, "Optimal daltonization by spectral shift for dichromatic vision," in *Color and Imaging Conference*, (Society for Imaging Science and Technology, 2012), pp. 302–308.

24. Color Imaging Lab, "Database of Ishihara test," [https://colorimaginglab.ugr.es/pages/Data#\\_doku\\_ishihara\\_spectral\\_database](https://colorimaginglab.ugr.es/pages/Data#_doku_ishihara_spectral_database).
25. S. M. Nascimento, F. P. Ferreira, and D. H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *J. Opt. Soc. Am. A* **19**(8), 1484–1490 (2002).
26. D. H. Foster, S. M. C. Nascimento, and K. Amano, "Database of rural and urban scenes," [https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral\\_images\\_of\\_natural\\_scenes\\_02.html](https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_02.html).
27. Columbia University, "CAVE database," [https://www.cs.columbia.edu/CAVE/databases/multispectral/real\\_and\\_fake/](https://www.cs.columbia.edu/CAVE/databases/multispectral/real_and_fake/).
28. N. Ohta and A. Robertson, *Colorimetry: fundamentals and applications* (John Wiley & Sons, 2006).
29. V. C. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm," *Vision Res.* **15**(2), 161–171 (1975).
30. S. L. Merbs and J. Nathans, "Absorption spectra of the hybrid pigments responsible for anomalous color vision," *Science* **258**(5081), 464–466 (1992).
31. J. L. Barbur and M. Rodriguez-Carmona, "Variability in normal and defective colour vision: Consequences for occupational environments," in *Colour design*, (Elsevier, 2012), pp. 24–82.
32. H. Kotera, H. Motomura, and T. Fumoto, "Recovery of fundamental spectrum from color signals," in *Color and Imaging Conference*, (Society for Imaging Science and Technology, 1996), pp. 141–144.
33. G. Wyszecki, "Psychophysical investigation of relationship between normal and abnormal trichromatic vision," *Farbe* **2**, 39 (1953).
34. J. B. Cohen and W. E. Kappauf, "Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks," *The Am. J. Psychol.* **95**(4), 537–564 (1982).
35. F. Ebner and M. D. Fairchild, "Development and testing of a color space (IPT) with improved hue uniformity," in *Color and imaging conference*, (Society for Imaging Science and Technology, 1998), pp. 8–13.
36. L. E. Lipetz, "Universal human visual pigment curves from psychophysical data," *Color Res. Appl.* **13**(5), 276–288 (1988).
37. M. D. Fairchild, *Color appearance models* (John Wiley & Sons, 2013).
38. M. F. Hassan and R. Paramesran, "Naturalness preserving image recoloring method for people with red–green deficiency," *Signal Process Image Commun* **57**, 126–133 (2017).
39. O. Fidaner, P. Lin, and N. Ozguven, "Fidaner recoloration method," [http://scien.stanford.edu/class/psych221/projects/05/ofidaner/project\\_report.pdf](http://scien.stanford.edu/class/psych221/projects/05/ofidaner/project_report.pdf).
40. H. Brettel, F. Vienot, and J. D. Mollon, "Computerized simulation of color appearance for dichromats," *J. Opt. Soc. Am. A* **14**(10), 2647–2655 (1997).
41. L. Zhang, Y. Shen, and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. on Image Process.* **23**(10), 4270–4281 (2014).
42. S. Le Moan, J. Preiss, and P. Urban, "Evaluating the multi-Scale iCID metric," in *Image quality and system performance XII*, (International Society for Optics and Photonics, 2015), pp. 939612.
43. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. on Image Process.* **15**(2), 430–444 (2006).
44. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013).
45. I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker, "Image-difference prediction: from grayscale to color," *IEEE Trans. on Image Process.* **22**(2), 435–446 (2013).
46. I. Lissner and P. Urban, "Toward a unified color space for perception-based image processing," *IEEE Trans. on Image Process.* **21**(3), 1153–1168 (2012).
47. D. J. Ketchen and C. L. Shook, "The application of cluster analysis in strategic management research: an analysis and critique," *Strat. Mgmt. J.* **17**(6), 441–458 (1996).