IEEE TRANSACTIONS ON IMAGE PROCESSING VOL: 24 NO: 3 YEAR 2015

Saliency-based color accessibility

Satohiro Tajima*,1,2,3 and Kazuteru Komine[3]

January 14, 2015

Abstract **Perception of color varies markedly between individuals because of differential expression of photopigments in retinal cones. However, it has been difficult to quantify the individual cognitive variation in colored scene and to predict its complex impacts on the behaviors. We developed a method for quantifying and visualizing information loss and gain resulting from individual differences in spectral sensitivity based on visual salience. We first modeled the visual salience for color-deficient observers, and found that the predicted losses and gains in local image salience derived from normal and color-blind models were correlated with the subjective judgment of image saliency in psychophysical experiments, i.e., saliency loss predicted reduced image preference in color-deficient observers. Moreover, saliency-guided image manipulations sufficiently compensated for individual differences in saliency. This visual saliency approach allows for quantification of information extracted from complex visual scenes and can be used as an image compensation to enhance visual accessibility by color-deficient individuals.**

Index Terms **visual saliency, color vision, individual difference.**

* Corresponding author; Email: satohiro.tajima@gmail.com

1 Department of Neuroscience, University of Geneva. 1 rue Michel-Servet, Genève, 1211, Switzerland.

2 Brain Science Institute, RIKEN. 2-1, Hirosawa, Wako, Saitama, 351-0198, Japan.

3 Science & Technology Research Laboratories, Japan Broadcasting Corporation. 1-10-11 Kinuta, Setagaya-ku Tokyo 157-8510, Japan.

1 Introduction

The exclusivity of subjective experience partly depends on individual differences in sensory perception. For example, congenital differences in the expression of color-sensitive photopigments alter individual's perception of complex visual scenes [1, 2]. Thus, differences with regard to visual perception exist among individuals. A majority of humans are trichromatic, i.e., different colors are encoded by the population-activity of long-wavelength (L)-, middle-wavelength (M)-, and short-wavelength (S)-sensitive cones. Observers lacking one class of photopigments are deficient in their ability to discriminate colors; i.e., they are blind to the difference between a certain pair of colors, such as red or green. This divergence in color vision is a major consequence of genetic polymorphisms [2] (Fig. 1a). There are currently few practical methods to medically compensate such color-vision deficient (CD) observers so that they have common trichromatic vision, although gene therapies under development may soon provide a cure [3]. An alternative strategy to enhance visual accessibility for CD individuals is the compensatory design of visual material with sufficient information to be perceived by all viewers, including CD observers [4, 5, 6, 7].

Several graphical techniques can partially mimic the visual experiences of CD individuals [8, 9] by replacing each set of indistinguishable colors by a single color chosen to be along one of the chromatic confusion lines of dichromats [10, 11, 12] (Fig. 1b). This procedure provides precise chromatic metamers for dichromats. However, interpreting the simulation results is problematic because the selection of representative colors is not uniquely determined but contains ambiguity that depends on the algorithm used. For example, various combinations of red, green, and yellow can be used to simulate the appearance of those colors in "red–green" dichromat vision. Moreover, it has been difficult to evaluate unstructured visual data, such as natural images, directly from these simulations. It is difficult to choose the best image parameters because the interpretation of a simula-

tion's output depends on subjective impressions by the viewer, and comparing subjective experiences like color perception among individuals is not straightforward. A possible approach to comparing such subjective experiences is mapping the high-dimensional perceptual content onto a common low-dimensional subspace and defining a perceptual difference metric as the distance measured in that subspace. A naïve way to do this is to measure the chromatic contrast (distance in a specific color space) in the simulated images, but the problem of the ambiguity in color mapping remains unsolved. In addition, a simple color-difference analysis cannot account for counterexamples where dichromatic vision is superior to normal trichromacy: dichromat observers sometimes have advantages at finding camouflaged objects (e.g, predator/prey in the natural environment) against a multicolored background [13, 14, 15, 16, 17], and its ecological function is capturing attention in the context of animal and human behaviors [16, 17, 18, 19].

In this study, we propose visual saliency [20, 21, 22, 23] as an alternative subspace in which to quantify perceptual differences (Fig. 1c). The concept of visual salience was first proposed in the context of cognitive science [20], and later specific computational implementations were proposed [22]. Saliency is widely used in studies on human visual attention, which is affected by the spatiotemporal context in visual stimulus; for example, it has been intensively reported that saliency models predict many properties of bottom-up visual attention in trichromatic human observers [23, 24, 25]. Although generally the term "saliency" can reflect various aspects of perceptual performance, including eye movements, in this study we will use this term to denote the visual conspicuity predicted from image features. As implied by previous studies [15], the notion of visual saliency potentially accounts for the complexity in the relationships between color vision and cognitive ability. However, there have been surprisingly few studies that have examined the saliency model and its behavioral implications for observers lacking trichromatic color vision, despite suggested importance of color information in object recognition by human [26, 27] and machine vision [28, 29]; for example, psychophysical study suggests that color information interacts with visual object representation at multiple processing levels [27]. In order to test whether and how much the proposed saliency analysis is relevant to human perception, we developed a psychophysical method to compare the perceptual content across observers in terms of visual salience, and the results suggested that a theoretically derived saliency-loss value ($\mathcal{L}$) is correlated with the difference in subjective perceptual saliency reported by normal and CD observers. Moreover, further experiments revealed that this difference in image preference could be ameliorated by manipulating the visual stimuli so that there would be a low saliency difference between trichromat and CD observers, thereby establishing a causal relationship between the quantified image salience and the perceptual judgments. This finding underscores the reliability of the present saliency-based approach for measuring perceptual divergence ressulting from variability in color vision and its potential for quantitatively relating color vision polymorphism and behavioral divergence. The present approach also provides a promising and practical strategy for the design of visual content accessible by observers with a range of spectral sensitivities, including those with color-vision deficiency.

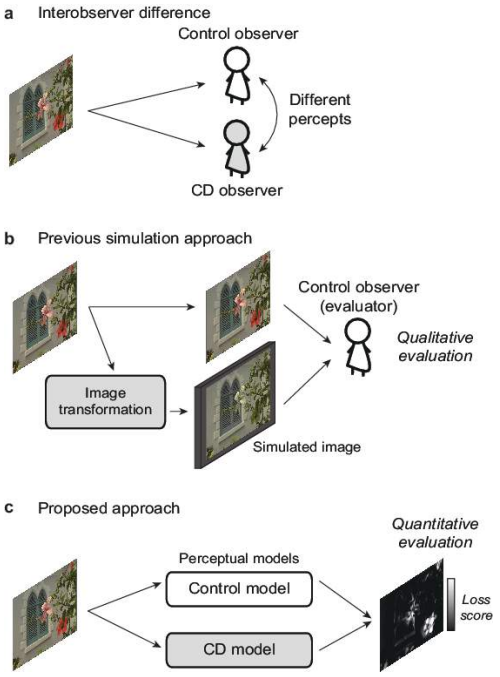

Figure 1: The considered problem and the proposed approach. (a) The problem. Two observers with different color visions have distinct perceptual experiences for the physically identical stimulus. (b) Previous approach based on color conversion. (c) The proposed approach based on visual salience.

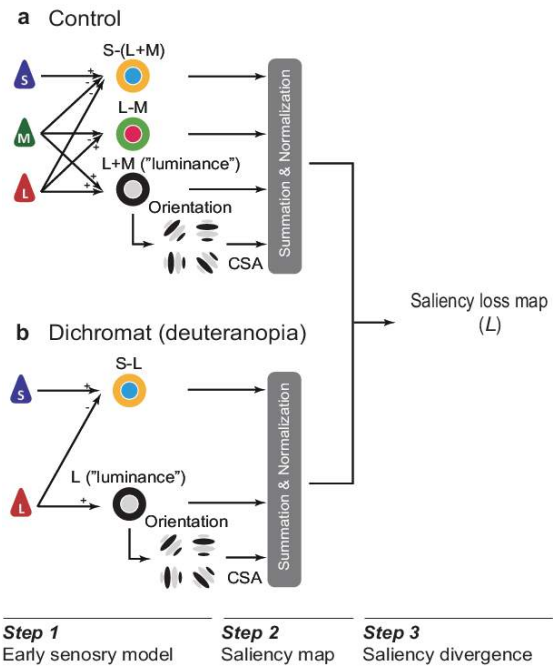

Figure 2: Computation of the saliency loss. (a) A model of saliency map for control observer having the common trichromatic vision. (b) Hypothetical computation of saliency maps from the responses of dichromat observers, where the figure shows the case of deuteranopia. CSA: center-surround antagonism.

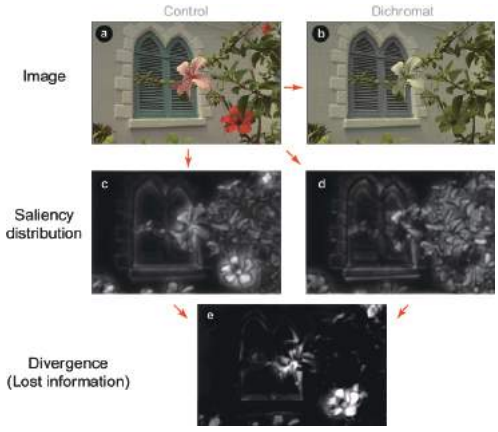

Figure 3: Computation of saliency divergence. (a) Original image. (b) Simulation of dichromatic vision. (c) Saliency map computed for a trichromatic observer. (d) Saliency map computed for a dichromatic observer. The arrows indicate the flow of processing; note that the dichromatic saliency map is directly derived from the original image instead of the simulation shown in (b). (e) The distribution of information loss visualized as the divergence between two saliency maps.

2 Methods

2.1 Model of visual information divergence

To compare the color perception between trichromatic and dichromatic observers when viewing complex visual scenes, we introduced the metric of saliency loss ($\mathcal{L}$). If we have two saliency maps that are derived from two different visual processing models (e.g., trichromacy and dichromacy), the structural divergence between the maps will ideally capture the characteristics of perceptual differences between the two observers. The saliency divergence was derived in the three steps described below: (1) models of the early sensory system, (2) derivation of saliency maps, and (3) quantification of saliency divergence. The first two steps can be applied to different visual processing models by changing their parameters.

2.1.1 Models of the early sensory system

Opponent color maps. First, we predicted the early sensory encoding of color, luminance, and orientation, according to models of the cone distribution and subsequent opponent-color processing (Figs. 2a and 2b). Note that the previously proposed simulation methods (e.g., [8, 9]) do not consider the color-opponency system but reconstruct the images on the basis of cone response (LMS) space to simulate the color confusions experienced by dichromat observers. To link the cone configuration and later visual processing stage, we extended the original model of visual salience [22] by introducing a more biologically plausible model of color opponency in the early visual system (Table. 1); here, the chromatic information is represented in the Derrington-Krauskopf-Lennie (DKL) color space, which reflects the preferences of retinal ganglion cells and lateral geniculate nucleus (LGN) neurons [30] and may describe the bottom-up attention resulting from early visual processing [31].

Table 1: Models of early visual system for the individual color vision types.

	Cones			Opponent signals		
	S	M	L	"blue–yellow"	"red–green"	luminance
Protanope	√	√	–	S–M	–	M
Deuteranope	√	–	√	S–L	–	L
Tritanope	–	√	√	–	L–M	L+M
Common	√	√	√	S–(L+M)	L–M	L+M

To date, there is no established color space model for dichromats or anomalous trichromats, although some studies suggest that "red" and "green" can be discriminated through variations in the optical density of pigments [32] or additional nonlinear computations based on existing cones [33]. In this study, we simply assumed that the color opponency of the missing cone type is absent (Fig 2b): that is, in the computation of the luminance and chromatic signals, an "L+M" signal was substituted with M for protanopia, and with L for deuteranopia. (e.g., for protanopia, the "luminance" and "blue-yellow" sensitive components were proportional to M and S/M, respectively). The L-M dimension was not computed for protanopia and deuteranopia, while the S dimension was not computed for tritanopia. We modeled opponent-color spaces for anomalous trichromats in the same way as the common trichromat but with different cone sensitivity functions [7]: we utilized the L (M) cone sensitivity functions for protanomaly (deuteranomaly) observers that is reported in [34]. Finally, we assumed that the visual stimuli are viewed on a display with sRGB color primaries, and used Hunt-Pointer-Estevez transformation for making the conversion from CIE XYZ to LMS space [35]. This is a parsimonious model that follows basic findings in the anatomical connectivity in the early visual systems and a previously verified model [7], without introducing any unreported physiological connections. The point of our model is to introduce the information loss at the level of opponent color coding; we also observed that the minor changes in the model, including the weighting parameters to computing opponent-color signals, did not affected the main results.

Orientation maps. The orientation maps were derived from the output of the luminance signal. The orientation filters were products of von Mises orientation selectivity [36] and log-Gaussian spatial frequency selectivity [37, 38], which well describe the neural receptive fields in the early visual cortex:

$$w_{ij}(\mathbf{k}) = e^{\kappa \cos(\angle \mathbf{k} - \theta_i)} e^{-\frac{1}{2}\left(\frac{\ln(\|\mathbf{k}\| - \phi_j)}{\sigma}\right)^2} \qquad (1)$$

where w_{ij} is the filter weight that peaked at the ith orientation and jth spatial frequency, (θ_i, ϕ_j). $\mathbf{k}$ denotes a wave vector in the spatial frequency space that was derived with a two-dimensional Fourier-transform, $\angle\mathbf{k}$ and $|\mathbf{k}|$ respectively represent its orientation and absolute value. The tuning parameters were set as $\kappa = 3$ and $\sigma = .5$, which roughly approximated the tuning property in the primary visual cortex [39]. Multiplying the above weight function in the spatial frequency space simulates convolution of image with a receptive field that is sensitive to local orientation within images. We mainly used four orientation bands (vertical, horizontal, and two diagonal orientations) and four spatial scales. We also explored the effect of varying the number of spatial scales in the additional analysis. Variation in other major parameters (increasing the number of the orientation bands, or eliminating the highest or the lowest spatial scale) did not yield any qualitative changes. We did not extract temporal features since we used only static images.

2.1.2 Derivation of saliency maps

Center-surround antagonism. Saliency maps (Figs. 3c, d) were constructed based on the aforementioned models of early sensory responses. The maps of different modalities (color, luminance, and orientation) were first independently processed by multi-scale center-surround antagonistic filters. First, the local features in the luminance, chromatic, and orientation maps were extracted with center-surround antagonism (CSA) filters. After pixel-wise square rectification of each modality map, CSA filters were applied on multiple spatial scales by decomposing the image. Again we used log-Gaussian-type band-pass filters, which can capture the scale-free property of natural images [40, 41]. The CSA filter was defined in the spatial frequency space as follows:

$$CSA_l(\mathbf{k}) = \exp\left(-\frac{1}{2}\left(\frac{\ln(\|\mathbf{k}\| - \psi_l)}{s}\right)^2\right) \qquad (2)$$

where ψ_l is the spatial frequency at which the lth filter has its peak weight. The peak spatial frequency of the finest band-pass filter, ψ_1, was matched to a visual angle of 15 cycles/deg (approximately equal to the half limit resolution of an observer with a visual acuity of 1.0). The lower spatial frequency bands had peak weights at the halves of the nearest higher frequency bands, i.e., $\psi_l = \psi_{l-1}/2$ $(l = 2, 3, 4, \cdots)$. The smallest spatial-pooling scale (the peak frequency of the finest band-pass filter) corresponds to pooling within a disc of diameter 0.07°. Multiplying these functions in the frequency space simulates the convolution of Mexican-hat-like filters across a feature map, which extracts local contrasts of the spatial configuration of the feature on multiple scale.

Saliency map. Next, individual CSA filter outputs were inverse-Fourier transformed and square-rectified, in order to produce feature maps $f_l(\mathbf{r})$ of pixel location $\mathbf{r}$ for different spatial scales l. Individual pixels in these within-scale feature maps were

summed up across all the spatial scales (L-M and S components were summed up to derive a color feature map; four orientation maps were summed up to derive an orientation feature map), before they were normalized (i.e., divided by the sum of the values within the whole image):

$$F(\mathbf{r};m) \equiv \sum_l f_l(\mathbf{r};m) / \left(\sum_{\mathbf{r}} \sum_l f_l(\mathbf{r};m) \right)^p \quad (3)$$

where the parameter p controlled the effectiveness of the normalization; we set $p = 1$ for the main analysis, roughly following Heeger-type normalization [42]. This normalization procedure mimics the neuronal gain control in the visual pathway [42]. The results were three feature maps for different modalities m (luminance, color, and orientation).

Finally, the luminance, color and orientation feature maps were further unified into a single saliency map by computing their weighted sum for each pixel: $S(\mathbf{r}) = \sum_m F(\mathbf{r};m)$. We set the weights of the summation to 1 of the luminance and chromatic maps and 0.1 for the orientation map for the main analysis. This unified saliency map was again normalized by the sum of values within the whole image, and this yielded the final saliency map at pixel $\mathbf{r}$ in image I for an observer O_i: $P(\mathbf{r}|O_i;I) = S(\mathbf{r}) / \sum_{\mathbf{r}} S(\mathbf{r})$, as a probability distribution. Note that the saliency map for a dichromat observer (Fig. 3d) is not identical to the map obtained by using the simulated dichromatic image (Fig. 3b) in the conventional trichromatic saliency analysis.

2.1.3 Quantification of saliency divergence

Finally, a spatial distribution of "saliency loss and gain" was obtained by comparing the two different saliency maps (Fig. 3e) between groups (normal vs. CD). We quantified the local information loss (or gain) value $\mathcal{L}$ for observer O_i compared with another arbitrary observer O_j as

$$\mathcal{L}(\mathbf{r};I) \equiv \ln P(\mathbf{r}|O_i;I) - \ln P(\mathbf{r}|O_j;I) \quad (4)$$

where ln denotes the natural logarithm. A negative $\mathcal{L}$ value means a saliency gain for observer O_i, and we refer it as a "saliency gain". Note that if we regard saliency value $P(\mathbf{r}|O_i)$ at pixel pixels $\mathbf{r}$ as the putative probability density of attentional attraction, the expected value of loss $\mathcal{L}$ over the whole image for a given observer O_i corresponds to the Kullback–Leibler divergence (an information-theoretic measure of the distributional difference, or *pseudo-distance*[43]) between the saliency maps for two observers:

$$\begin{aligned} &\mathcal{D}(\mathbf{r};I) \quad (5) \\ &\equiv D_{KL}[P(\mathbf{r}|O_i;I)||P(\mathbf{r}|O_j;I)] \\ &= E[\mathcal{L}(\mathbf{r};I)|O_i] \\ &= \sum_r P(\mathbf{r}|O_i;I) \left\{ \ln P(\mathbf{r}|O_i;I) - \ln P(\mathbf{r}|O_j;I) \right\} \end{aligned}$$

This can be interpreted as the distance between for the two saliency maps, measured from observer O_i to observer O_j. Note that the above measure takes zero value only if $P(\mathbf{r}|O_i;I) = P(\mathbf{r}|O_j;I)$ for every $\mathbf{r}$.

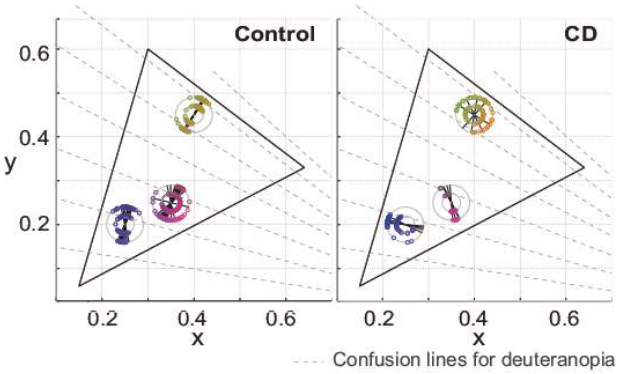

Figure 4: Results of color-vision characterization with color matching task (selecting the closest matching color to the reference color) in the control and color-vision-deficient (CD) groups. The colored markers indicate the points that the participant judged as the same the reference color. The gray circles depicts the manifolds from which the participants selected the matched colors. The black lines indicates the angular mean of the selected colors for each individual. The data for all control and CD participants are superimposed.

2.2 Psychophysical experiments

2.2.1 Experiment 1: Dichromat simulation

This experiment was for testing whether and how the saliency-based model can be related to the interobserver differences in the characteristics of color vision.

Participants Five CD participants (males, 22-34 years old) and 18 control participants (12 females and 6 males, 18-40 years old) with normal color vision judged the subjective saliency of the presented images. The study was approved by the Ethics Committee of the Science & Technology Research Laboratories, Japan Broadcasting Corporation, and performed

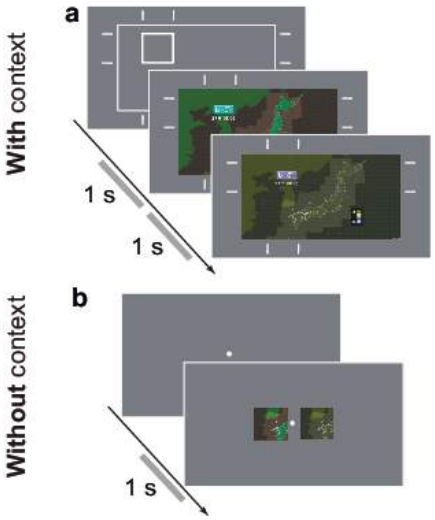

Figure 5: The visual stimuli for Experiment 1. The participants compared the original image vs. the corresponding dichromat simulation (DS), which were presented (a) with or (b) without spatial context.

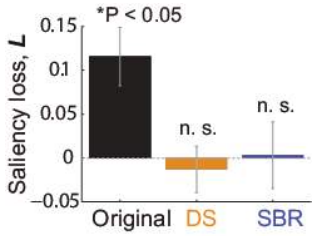

Figure 6: The information loss values in the original and SBR images. Error bars indicate the standard error of the mean.

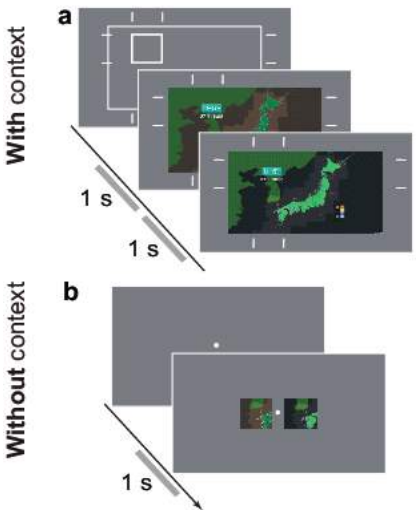

Figure 7: The visual stimuli for Experiment 2. The participants compared the original image vs. the corresponding image after modification by saliency-based recoloring (SBR), which were (a) with or (b) without spatial context.

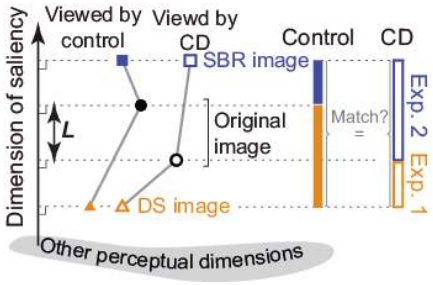

Figure 8: Schematic illustration of the experimental design.

according to the Declaration of Helsinki. All subjects gave written informed consent before taking part in the study. The visual properties of the participants were doubly confirmed by their taking the Ishihara test [44], which is used for discriminating congenital color deficiency from normal vision, and performing an additional color matching task that was modified from [45].

The Ishihara test score suggested that five had CD. The CD participants showed responses that were consistent except for a single participant who responded differently from the other four and from the response pattern described in the Ishihara chart. This participant was excluded from the main analysis of the experiment to ensure relative uniformity of visual abilities within each group. However, we also confirmed that the general pattern and statistical significance of the main results were maintained when we included this participant in the analysis.

The results of the color matching task also confirmed the within-group consistency and between-group dissociation of sensitive dimensions in the chromatic coordinate, as well as that the confusing color for four deuteranope participants were consistent with what was reported previously [10] (Fig. 4). We should emphasize here that neither the Ishihara test nor the color matching task was not for a precise discrimination between color-deficiency types or dichromacy and anomalous trichromacy, but were for roughly confirming dissociation between the two groups and for showing a general within-group con-

sistency of the trend in terms of the ability to discriminate colors in the chromatic diagram.

Stimuli Colored illustrations were compared with the corresponding modified versions generated by *dichromat simulation* (DS) (Fig. 5a and 5b). The original images were acquired from a database of pictures used in television programs. The corresponding DS image was an approximation of deuteranope vision created using a method proposed by Brettel et al. [9]. The target of judgment was restricted to a specified region of the image (size of $6.75^\circ \times 6.75^\circ$) presented within the whole image ("with context", Fig. 5a) or in a cropped form ("without context", Fig. 5b). The purpose of removing the spatial context was to eliminate the possible effect of semantic stimulus interpretation on the judgment. Six different regions within each image pair were judged four times by each participant. The computation of the hypothetical saliency loss predicted that the original images would be perceived as less salient when viewed by CD observers than when viewed by control observers, while there should be no significant observer-dependent loss for DS images (Fig. 6).

Procedures Experiments were conducted in a dark room. Stimuli were presented on gamma-corrected LCD monitors (EIZO, colorEdge CG276 or CG246). The with-context stimulus subtended $30^\circ \times 16.8^\circ$ of visual angle and the without-context image subtended $6.75^\circ \times 6.75^\circ$ when participants sat 90 cm from the monitor. Maximum luminance was 120 cd/m^2 and minimum light levels were below 0.01 cd/m^2. All presentation and data analyses were performed using the MATLAB (Mathworks, Natick, MA) programming environment and the Psychtoolbox extension ([46, 47]).

In all experiments and conditions, each trial started with the presentation of a central fixation disk for 200 ms before stimulus onset. The participants judged whether the original or the DS-modified image appeared clearer in a two-alternative forced-choice task. We asked our participants to judge the "clarity" of images, by describing "how easily the visual information is extracted from images", instead of directly judging visual salience because naïve participants might find it hard to understand the notion of "saliency". In the with-context condition, full images (original and DS-modified) were presented in sequence at the same central position of the screen for 1 s at 300 ms intervals. The temporal order of the two images was randomized over the trials. In each trial, the target image region was indicated by a box presented before stimulus presentation and by lines during stimulus presentation (Fig. 5a). In the with-context condition, the cropped images were simultaneously presented to the left and right of the fixation point (Fig. 5b) with the inner edges 3˚ from the fixation point. The spatial positions of the two images (original or DS) were randomized over the trials. In the with- and without-context conditions, each stimulus was presented for 1 s and followed by a brief random noise mask lasting 200 ms to eliminate the sensory afterimages.

Data analysis The mean responses are expressed as the probability of selecting the original image as more salient than the DS image. In the correlation analysis, we computed the correlation between the stimulus-wise judgment and the local average of saliency loss within the DS stimulus: $\ln P(\mathbf{r}|O_{Control}; I_{Original}) - \ln P(\mathbf{r}|O_{Control}; I_{DS})$. In the differential analysis, we computed the stimulus-wise difference in judgment by the control and CD participants. The stimulus-wise judgments were also correlated to the average saliency loss between the control and CD observers, $\mathcal{L}(\mathbf{r}; I_{Original}) = \ln P(\mathbf{r}|O_{Control}; I_{Original}) - \ln P(\mathbf{r}|O_{CD}; I_{Original})$. Note that the loss and the gain can cancel out each other when they are averaged across a wide visual field. To avoid this effect, our experiments were designed to measure local properties by instructing the participants to judge relatively small regions within stimuli, rather than global impressions of the whole images, by limiting the image regions to be judged by the participants (Figs. 5a and 5a). Finally, we also tested the correlation between the same experimental data and predictions by other alternative methods based on color contrast, where the model predictions were provided with local spatial averages of pixel-based square distances between original and of DS images represented in LMS, DKL, or RGB spaces.

2.2.2 Experiment 2: Saliency-based recoloring

This experiment was for testing whether the saliency-guided manipulation of images can compensate for perceptual differences between control and CD participants. The subjects, images of sources, and experimental procedure were the same as in Experiment 1 except for that we used images that were modified by *saliency-based recoloring* (SBR; Fig. 7a and 7b) instead of DS images. These SBR images were obtained through a recoloring procedure as follows: first, the RGB image was converted to the DKL representation. Second, L–M signals superimposed upon the

(L+M) and S–(L+M) signals, $I^{new}_{L+M} = (1-w)I^{old}_{L+M} + wI^{old}_{L-M}$, $I^{new}_{S-(L+M)} = (1-w)I^{old}_{S-(L+M)} + wI^{old}_{L-M}$, where $I^{new}_{\cdot}$ and $I^{old}_{\cdot}$ respectively represent the recolored and original image signals (in the DKL color space) that correspond to the channels shown in the subscripts, and w is the linear mixing weight. To obtain SBR image, the weight w was determined to minimize the information loss described in Eq. 4, i.e., $I_{SBR} \equiv \arg\min_I \mathcal{D}(\mathbf{r}; I)$, where $\mathcal{D}(\mathbf{r}; I) = \sum_r P(\mathbf{r}|O_i; I) \{\ln P(\mathbf{r}|O_i; I) - \ln P(\mathbf{r}|O_j; I)\}$. In this study, the minimization was achieved through hand-tuning of w with monitoring of the loss value in realtime although the automation of this procedure is straightforward. In addition, the post-hoc analysis confirmed effective reduction in the saliency loss at each local image region (Fig. 6). Finally, we reverse mapped I_{SBR} from the DKL space to RGB space. The solution of this minimization problem is not unique; in this study, as a rough optimization, we first superimposed L-M signals upon the luminance and blue–yellow signals at various random strengths and selected the optimal image by referring to the information loss calculated for each version of image. After this procedure, it was confirmed that the SBR images had smaller information loss (L) values that were distributed around zero, while the values for the original images were significantly larger than zero ($t_{23} = 2.50$, $P < 0.05$; Fig. 6). We used four of the six images of the sources from Experiment 1. In addition, similar to Experiment 1, we tested six spatial regions for each of the four source images (i.e., $6 \times 4 = 24$ local patches in total). In the integrated analysis of Experiments 1 and 2 (corresponding to Figs. 12a and 12b, right columns), we analyzed the 24 images common to both experiments by two-way ANOVA with stimulus type (DS vs. SBR − DS or SBR vs. SBR − DS) and context (with- and without-context) as main factors.

2.2.3 Relationship between the two experiments

The relationship between Experiments 1 and 2 is schematically expressed in the perceptual space of Fig. 8. Theoretically, SBR images should yield similar saliency maps for the control and "red–green" CD observers (as indicated by the blue squares in Fig. 8) because they do not contain information solely encoded by the L–M dimension. If this is the case, the saliency-relevant distance between the SBR and the original image should be longer for the CD observer than that for the control observer (the open and filled blue bars). In addition, the interobserver differences measured in Experiments 1 and 2 are expected to approximately nullify each other. Thus, the interobserver differences in perception should vanish when the individual perceptual distances measured in the two experiments are summed (Fig. 8, blue bar + orange bar).

These facts are schematically expressed in terms of the saliency-relevant perceptual dimension (subspaces) and the other dimensions (Fig. 8). The original images have different values depending on the color vision (normal or deuteranopic) of the observer in the saliency dimension (as indicated by the open and solid black circles). However, the DS images have almost the same saliency for both observer groups (orange triangles). The latter result is because of the lack of the L-M component. Thus, both CD and control observers should have access to the same visual information, and we can reasonably assume that DS images provide baseline saliency that is common to both sets of observers. Moreover, these DS images enable us to quantify the perceptual differences reported by CD and control observers (Fig. 8, filled and open orange bars). DS images are metamers for pure dichromat observers [8, 9]; therefore, the original and DS images are also expected to overlap in the saliency domain. For CD observers with anomalous trichromacy (as in the present experiment), the distance between the original and DS images should have a non-zero saliency-defined value, although it is expected to be smaller than that for control observers.

3 Results

3.1 Saliency losses and gains in complex visual scenes

Figures 9 presents the saliency divergence for a variety of visual stimuli. For example, Fig. 9b shows the results for a representative plates from the Ishihara test chart [44], which is used for discriminating congenital "red–green" color deficiency from normal vision. The chart is characterized not only by the use of indistinguishable colors but also by the spatial configuration of the dot pattern. The model yields the plausible result that protanopes (lacking L cones) and deuteranopes (lacking M cones) lose saliency in the region corresponding to the green figure "7," while tritanopes (lacking S cones) do not. In contrast, for more naturalistic stimuli (Figs. 9a, c,

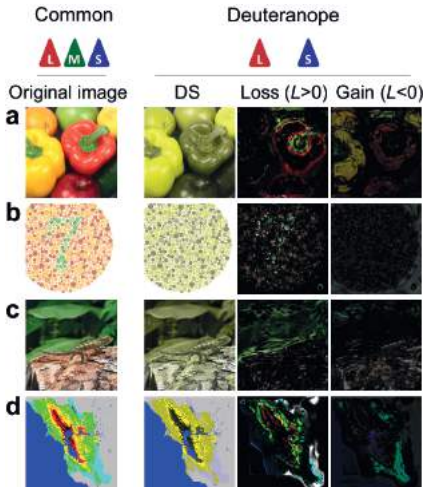

Figure 9: Examples of saliency losses and gains in Deuteranope observer for naturalistic and artificial visual stimuli. (a) Fruits and vegetables, (b) an Ishihara chart (adapted from "Ishihara Test for Color Deficiency" reprinted with permission from the Public Interest Incorporated Foundation, Isshinkai, Tokyo, Japan), (c) wild camouflage by a chameleon in Madagascar, and (d) a seismic hazard map of the northern Hayward Fault Zone. For each dichromat type, the figure shows DS image and the highlighted image indicating the saliency decrease (with positive "loss" values, $\mathcal{L} > 0$) or increase (with negative "loss" values, $\mathcal{L} < 0$, or "gain") compared to the normal trichromatic vision. The exactly neutral situation (L=0) is mathematically possible, but rarely happens in general images; nonzero L values fill the most part of each image, as we see in these figures. Note that the color of the figure is not necessarily identical to the original because of the reproduction process, and the results of the simulation (e.g., the yellowish parts in the figure) is not intended to show the actually "perceived" colors in dichromats.

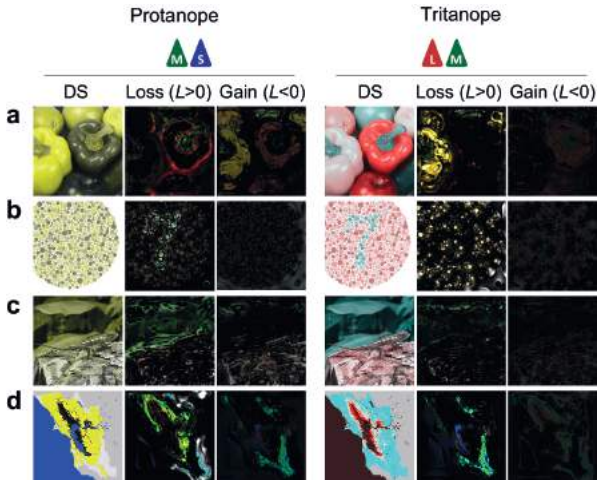

Figure 10: Examples of saliency losses and gains in Protanope and Tritanope observers. Panels a–d correspond to those in Fig. 9.

and d), the model predicts that some areas in the scene are significantly more salient for dichromats than for trichromats because of the normalization process in the saliency computation. The increased saliency provides a computational explanation for the reported advantage of dichromat animals in capturing camouflaged prey in the wild [48, 49, 18, 50] (e.g., 9c). The simple color-contrast model without normalization cannot account for such phenomena (see Discussion).

3.1.1 Difference and correlation in saliency loss across color-vision types

In Fig. 9, the model not only differentiates the saliency map for tritanopes from the other two dichromatic visions, but also yields slightly different predictions between protanopes and deuteranopes for some stimuli (e.g., Figs. 9a and 10a). This is because the difference in cone combination (i.e., L and S, or M and S) can lead to different luminance signals, in addition to the primary effect of lacking an L-M signal. This fact indicates that the current approach is potentially able to differentiate perceptions resulting from different types of color deficiency. However, it should be noted such precise predictions are possible only after we have detailed physiological data concerning how cones are directly and indirectly connected to the subsequent visual neurons in each color-vision type, which is not established at this stage. We will not discuss how well the saliency-based approach can predict the precise difference between protanopia and deuteranopia, or between deuteranopia and anomalous trichromacy, because we need much more detailed physiological information to construct realistic models for such aims.

Instead, in this study, we propose that the current model can nevertheless provide a practical approximation of saliency loss for a broad class of so-called "red-green" color vision deficiency, in a rough but still practically sense for the following reasons. As demonstrated in Figs.9 and 10, overall, the protanopic and deuteranopic visions yielded very close results in terms of saliency loss. This is because (1) the major cause for the saliency reduction is the lack in the L-M opponent signal, and (2) the signals conveyed by L-cones and M-cones are highly correlated for the most stimuli encountered in daily life [51]. This means that, for example, the deuteranopic model can also

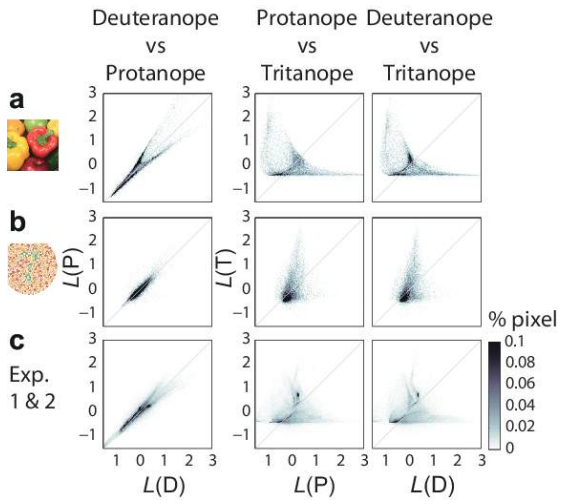

Figure 11: Correlated saliency loss between protanope and deuteranope. The local saliency loss computed with different cone combinations are plotted as 2D histograms. ($\mathcal{L}$(P): protanope; $\mathcal{L}$(D): deuteranope; $\mathcal{L}$(T): tritanope). Left column: protanope vs deuteranope; Middle column: protanope vs tritanop; Right column: deuteranope vs tritanope. (a–b) Corresponding results for the images shown in Figs. 9a and b. (c) Results for all stimuli used in the psychophysical experiments.

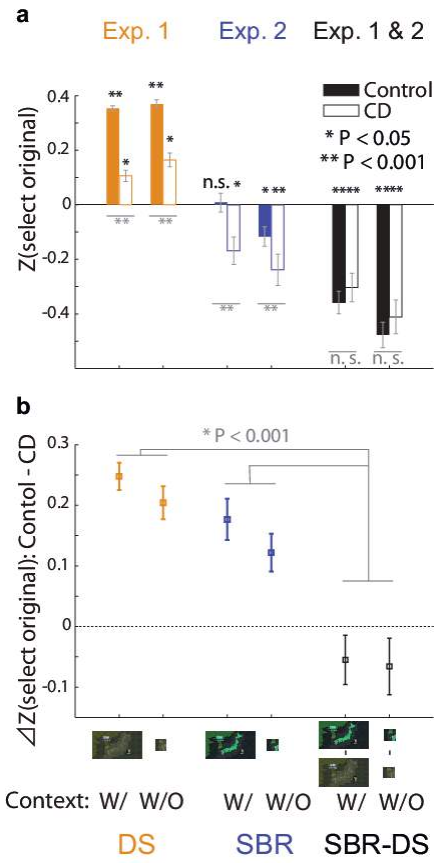

Figure 12: Results of the psychophysical experiments (1 and 2). (a) The probability (as expressed by Z-score) of selecting the original image as more salient for both the with- and without-context conditions. (b) The differences in Z-scores between the control and CD participants as presented in (a). The error bars in (a) and (b) indicate stimulus-to-stimulus standard errors.

provide reasonable first-order approximations also for protanopia, in terms of saliency for the majority of stimuli. Indeed, the model tends to make strongly correlated predictions for those two color-vision types (Fig. 11, left column), whereas it is not the case for comparisons between protanope and tritanope, or between deuteranope and tritanope (Fig. 11, middle and right columns). The same discussion applies to L- or M-cone related anomalous trichromacies, and it is conjectured that also those types of anomalous trichromacies can also be characterized by saliency losses with similar spatial patterns to deuteranopia, but with reduced magnitudes of losses and gains. Again, we note that detailed information acquired through future physiological investigations will be needed in order to determine more precise differences among these color-vision types.

3.2 Saliency-based prediction and manipulation of interobserver perceptual difference

3.2.1 Saliency loss predicts the variability in perception

In Experiment 1, the comparison of the average saliency judgments of the control and CD observers (Fig. 12a) was consistent with predictions of the model. The control participants demonstrated a strong tendency to judge the DS images as less salient than the original images, although this bias was more moderate in CD observers (with context: $t_{35} = 10.9$, $P < 0.001$; without context: $t_{35} = 7.38$, $P < 0.001$; two-tailed t-test with paired samples). These results support the plausibility of the deuteranope simula-

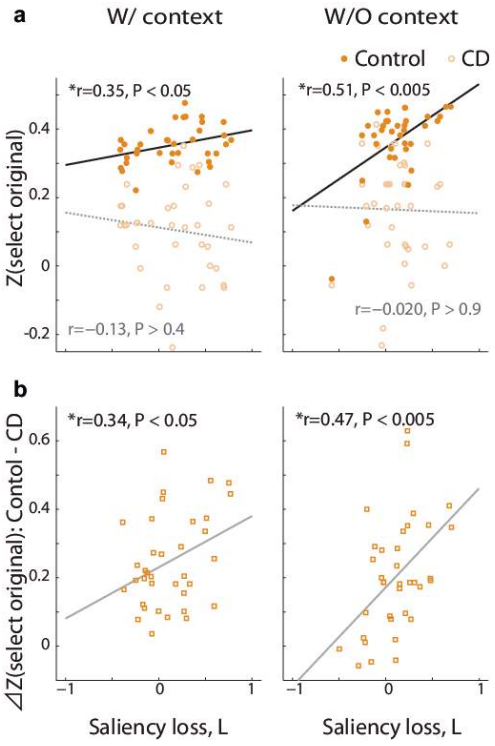

Figure 13: Correlation between psychophysical judgments and model predictions for 36 visual stimuli used in Experiment 1. (a) Correlation between the judgment Z-scores and the saliency-based information loss for individual stimuli. (b) Correlation between the difference in the judgment Z-scores (control − CD) and interobserver information loss based on the proposed salience method (Fig. 4). The lines in (a) and (b) indicate the results of linear regression.

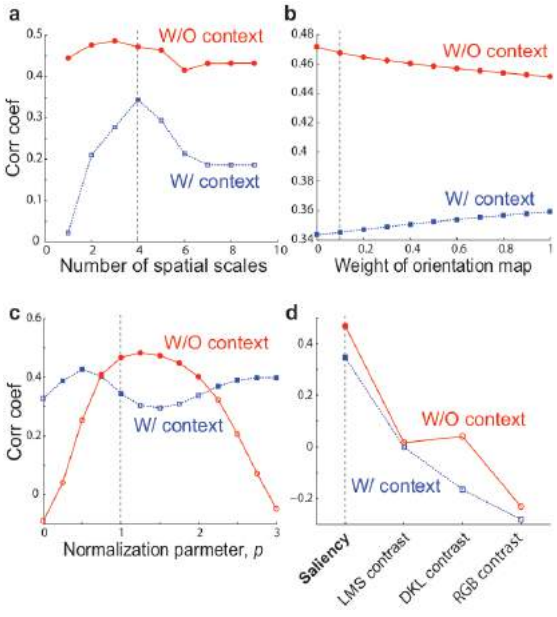

Figure 14: Parameter (a–c) and model (d) dependency of predictive accuracy for human judgments. The filled markers indicate the statistical significance (P<0.05, Pearson correlation test). Note that the vertical axes are differently scaled for individual panels.

tion and the saliency loss model for revealing individual differences in color perception.

It can be argued that the interobserver difference in the saliency judgment reflects some trivial response bias because of the unfamiliarity of DS images to the control participants compared with the CD participants. However, the theoretically predicted saliency loss positively correlated with the probability of control participants judging the original image as more salient, while no significant correlation was observed for the CD participants (Fig. 13a). This result supports the idea that the saliency judgment is determined by the perceived saliency because the correlation should not differ between groups if it is affected only by a constant response bias. In addition, we cannot rule out an additional contribution of a response bias (or some other confound) because there was also an intergroup difference for stimuli with saliency loss near zero. Moreover, we observed that the difference in judgment (control − CD) correlated with the hypothetical saliency loss (Fig. 13b). This correlation between the model prediction and the difference in the responses of the control and CD observers is consistent with the idea that the image saliency strongly correlates with the perceptual judgment. Although there were small negative correlations, they were not statistically significant and could be interpreted as the effect of noise in the judgment.

Model prediction performances were sensitive to change in the number of spatial-pooling scales and the normalization parameter (Fig. 14a, c), suggesting the importance of these parameters in describing the perceptual experiences by human observers. Other major model parameters, including the mixing weights of the orientation maps (Fig. 14b), had little effect on the model performance. Finally, we did not observe any significant correlation between the same psychophysical-judgments data and other alternative methods based on color contrast in LMS, DKL, or RGB spaces (Fig. 14d).

3.2.2 Saliency-guided image manipulation mitigates perceptual differences

Figure 12a (middle) shows the results of Experiment 2. As predicted, the SBR images were judged as more

salient by CD participants than by control participants (with context: $t_{23} = 5.08$, $P < 0.001$; without context: $t_{23} = 3.83$, $P < 0.001$; two-tailed t-test with paired samples). Notably, we also confirmed the hypothesized cancellation of perceptual saliency difference thorough SBR (see Seciton 2.2.3) as follows: the statistical significance of the interobserver difference vanished when the results of Experiments 1 and 2 were combined (Figure 12a, right; with context: $t_{23} = -1.32$, $P = 0.200$; without context: $t_{23} = -1.38$, $P = 0.180$; two-tailed t-test with paired samples). Thus, while we found significant interobserver perceptual differences in both Experiments 1 and 2, the combined interobserver differences were significantly smaller than those for either experiment alone (Experiment 1 [$F_{1,120} = 71.61$, $P < 0.001$] or in Experiment 2 [$F_{1,96} = 28.27$, $P < 0.001$]; one-way ANOVA). This is a reasonable result if it is assumed that the Z-scores obtained in these experiments reflect the linearized distances of visual stimuli in the perceptual space (see also Fig. 8). These results further support the view that participants judged saliency based on the saliency-relevant perceptual dimension and that matching the observer-dependent saliency maps can compensate for individual differences in perception.

4 Discussion

We proposed a new approach for quantifying and presenting interobserver differences in perceptual experience because of individual variations in spectral sensitivity . The two psychophysical experiments showed the plausibility of the proposed approach. These experiments revealed a strong statistical relationship between saliency loss or gain and the subjective judgment of saliency. Furthermore, we found that saliency-guided image manipulations were sufficient to compensate for interobserver differences in saliency. Thus, matching visual saliency is a promising strategy for ensuring that visual information is accessible to individuals with deficient color vision.

4.1 Multiscale representation and normalization

The unique aspect of the present approach is the quantification of interobserver differences in the salience domain. The saliency-based approach is distinguished from naive color-error analysis between original and DS images in two ways. First, the current approach integrates visual information with multiple spatial resolutions and features that reflect the early cortical representation of visual stimuli. The dependence of model predictability on the number of spatial scales to be integrated (Fig. 14a) suggests the importance of multiresolutional expression. Second, the proposed saliency computation includes a normalization process that predicts an increase in visual saliency for dichromatic observers under certain conditions. In Experiment 1, we observed that DS images with increased saliency for dichromatic observers were counterintuitively judged as more salient by these subjects in the without-context condition (Fig. 13b, right). Subtracting the L-M chromatic signals indicated that the perceptual saliency was increased for these observers; this result is accounted for by a mechanism that is intuitively similar to an image enhancement such as histogram equalization. In addition, the experiments varying the model parameters indicated that the normalization in Eq. 3 played an important role in terms of the model performance, at least in the without-context condition (Fig. 14c). With-context condition showed a more complex parameter dependency, which may reflect the contaminating effect of local features and higher-order cognitive processes.

We used "with-context" and "without-context" conditions to test how much the global visual context affects the judgments. There were similarities and difference between these two conditions. The qualitative trend was shared among the two conditions (Figs. 12 and 13), suggesting that the basic relationship between the saliency model and the human perception is robust to the changes in the spatial context. In addition, varying some of the parameters had different effects on the model predictability between the two conditions (Fig. 14). These differences can be reasonably interpreted from the viewpoint of the visual processing and stimulus characteristics. For example, the scale of spatial pooling is influences in the "with-context" condition but not in the "without-context"; this is not surprising because the stimulus was localized within a small visual field in the "without-context" condition, and variation in spatial pooling scale does not have a large effect on the model output in that condition. On the other hand, the normalization had more of an effect in the "without-context" condition than in the "with-context" condition; this might reflect the fact that the with-context stimulus had visual signals varying across spatial scales and it makes the sum across the spatial scale (the denominator in Eq. (3)) less affected by the stimulus.

It should be noted that we focused on local visual salience rather than holistic impressions on images. As mentioned before, although the loss and gain can compensate each other when they are aver-

aged across a wide visual field, our experiments were designed to quantify the local properties by instructing the participants to judge relatively small regions within stimuli, rather than global impressions of the whole images. In the data analysis, we related those psychophysical judgments with the local average of salience loss. Therefore, we believe that our experimental results appropriately supports our the model predictions.

4.2 Relevance to other methods

Some studies have discussed approaches for evaluating perceptual divergence among observers with different spectral sensitivities, e.g., by computing the judgment error due to the loss of contrast between original and dichromat-simulated images in specific color spaces [6, 5]. Although such methods are useful for a rough evaluation of color contrast in the image, they fail to capture many aspects of color vision in realistic situations. For example, suppose that the L-M signal was spatially correlated to the luminance signal: in this case, there is little loss in terms of visual salience between normal trichromacy and L- or M-cone related color dichromacy (or anomalous trichromacy), whereas the classical color contrast analysis (without normalization) would predict a significant reduction in information due to the color deficiency.

As another example, the simple chromatic-difference analysis cannot account for the complexity of the aforementioned counterintuitive cases where dichromats can be superior to normal trichromatic observers: a line of studies suggests that dichromats are better at finding camouflaged objects against a multicolored background [13, 14, 15, 16, 17], a perceptual ability that is critical for foraging, hunting, and avoiding predators in the wild [49, 18, 50]. The present saliency-based model with normalization process naturally accounts for such a possibility in which dichromacy becomes advantageous. At the single feature level, a previous study suggested that the normalization process can facilitate signal detection when combined with suppression of another specific signal [52]. On the other hand, the facilitation of saliency can be a realization of the "masking effect," which was conceptually hypothesized by Judd [14], by generalizing the relatively new concept of the neural activity normalization [42, 53] at a level of multiple feature maps.

It should be noted that we modeled the visual saliency specialized for the dichromatic observer (Fig. 2b) and derived it directly from the original image (Fig. 3a). This is not equivalent to naive application of the conventional trichromatic saliency model to DS or decolorized images [54] because the latter can be contaminated by the L–M cone contrast signals in the simulated image. For example, when the model−data correlation (Fig. 12d) was recomputed based on this naïve model, the correlation deteriorated (with-context condition: $r = 0.25$, $P = 0.12$; without-context condition: $r = 0.31$, $P = 0.058$; Pearson correlation test).

4.3 Limitations

A major limitation of the current saliency-based approach is that it does not account for the effects of semantic context on visual perception. Extraction of visual information is guided not only by low-level saliency (bottom-up attentional processes) but also by knowledge of the world, particularly when viewing familiar natural scenes. The contribution of semantic knowledge is consistent with the effect of context; there was a moderate decrease in model predictability for the with-context condition compared with the without-context condition (Figs. 13a and 13b), suggesting that additional information unrelated to salience per se can influence perceptual judgment. Nevertheless, we emphasize the importance of saliency as a critical determinant of bottom-up attention, the first bottleneck through which humans recognize the visual world. This "attentional metamerism" is a promising framework for discussing the universality of visual information, particularly when no common chromatic metamer exists because of interobserver differences in color sensitivity. These points could be elucidated by future research that directly examine the effects of color vision difference on the attention (and eye movement).

In the experiments, we characterized participants' ability at (red-green) color discrimination by using the Ishihara test and a color matching task. The current characterization is not sufficient for modeling precise differences among protanopia, deuteranopia, and anomalous trichromacies. In the future, experiments with a larger number of participants and more accurate diagnoses, e.g., based on the results of anomaloscope test or to analyze opsin genes, should be coducted to examine model predictability for the different types of color vision. Nevertheless, the present results suggest that our model works at least as a first-order approximation for observers who are less sensitive along the L-M ("red-green") dimension, indicating a practical way of improving image for them.

4.4 Practical implications

The proposed saliency-based method can be utilized for visual design to avoid unintended asymmetry in the information received by observers. Intriguingly, SBR images tended to be favored by both control and CD participants (Fig. 12a), indicating that the saliency-based approach can effectively enhance visual clarity for both control and color-deficient observers. The methodological framework for quantifying loss of visual information according to quantitative criteria is expected to enhance the effectiveness of a wide variety of visual stimuli. For example, in the process of creating information graphics (e.g., Fig. 9d), saliency analysis could be applied to determine whether some observers lose information at any given time point. In addition, the saliency-based approach has the potential to provide a common platform for accessibility assessments that can account for not only differences in color vision but also deficits in other aspects of vision, such as low acuity because of cataracts or macular degeneration and dysfunction of other sensory modalities, such as age-related degradation of auditory sensitivity.

Acknowledgments

We thank all the participants for their cooperation during the experiments. We also thank I. Kuriki, C. I. Tajima, and K. Uchikawa for discussions, and C. Hiramatsu for comments on the earlier version of manuscript. We thank M. Sawa for comments on the use of the Ishihara chart, and Japan Broadcasting Corporation (NHK) Nagano Station for support on the preliminary setup.

References

[1] J. Neitz, M. Neitz, and P. M. Kainz, "Visual pigment gene structure and the severity of color vision defects," *Science*, vol. 274, pp. 801–4, Nov. 1996.

[2] J. Neitz and M. Neitz, "The genetics of normal and defective color vision," *Vision Research*, vol. 51, pp. 633–51, Apr. 2011.

[3] K. Mancuso, W. W. Hauswirth, Q. Li, T. B. Connor, J. A. Kuchenbecker, M. C. Mauck, J. Neitz, and M. Neitz, "Gene therapy for red-green colour blindness in adult primates.," *Nature*, vol. 461, pp. 784–7, Oct. 2009.

[4] M. Ichikawa, K. Tanaka, and S. Kondo, "Web-Page Color Modification for Barrier-Free Color Vision with Genetic Algorithm," *Lecture Notes in Computer Science*, vol. 2724, pp. 2134–2146, 2003.

[5] K. Rasche, R. Geist, and J. Westall, "Reproduction of Color Images for Monochromats and Dichromats," *IEEE Computer Graphics and Applications*, vol. 25, no. 3, pp. 22–30, 2005.

[6] J.-b. Huang, S.-y. Wu, and C.-s. Chen, "Enhancing Color Representation for the Color Vision Impaired," in *Workshop on Computer Vision Applications for the Visually Impaired*, 2008.

[7] G. M. Machado, M. M. Oliveira, and L. A. F. Fernandes, "A Physiologically-based Model for Simulation of Color Vision Deficiency," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1291–1298, 2009.

[8] F. Viénot, H. Brettel, L. Ott, M. Ben, and J. D. Mollon, "What do colour-blind people see?," *Nature*, vol. 376, no. 6536, pp. 127–128, 1995.

[9] H. Brettel, F. Viénot, and J. D. Mollon, "Computerized simulation of color appearance for dichromats.," *Journal of the Optical Society of America. A, Optics*, vol. 14, pp. 2647–55, Oct. 1997.

[10] F. H. G. Pitt, "Characteristics of dichromatic vision," in *Medical Research Council Special Report Series*, vol. 200, p. 200, London: His Majesty's Stationery Office, 1935.

[11] L. C. Thomson and W. D. Wright, "The convergence of the tritanopic confusion loci and the derivation of the fundamental response functions," *Journal of the Optical Society of America*, vol. 43, pp. 890–4, Oct. 1953.

[12] D. Farnsworth, "Tritanomalous vision as a threshold function," *Die Farbe*, vol. 4, pp. 185–197, 1955.

[13] A. Anon, "Colour-blindness and camouflage," *Nature*, vol. 146, p. 226, 1940.

[14] D. B. Judd, "Color blindness and the detection of camouflage," *Science*, vol. 97, no. 2529, pp. 544–546, 1943.

[15] M. J. Morgan, A. Adam, and J. D. Mollon, "Dichromats detect colour-camouflaged objects that are not detected by trichromats," *Proceedings of the Royal Society of London B*, vol. 248, pp. 291–295, 1992.

[16] A. Saito, A. Mikami, S. Kawamura, Y. Ueno, C. Hiramatsu, K. Widayati, B. Suryobroto, M. Teramoto, Y. Mori, K. Nagano, K. Fujita, H. Kuroshima, and T. Hasegawa, "Advantage of dichromats over trichromats in discrimination of color-camouflaged stimuli in nonhuman primates," *American Journal of Primatology*, vol. 67, no. 4, pp. 425–436, 2005.

[17] A. Saito, A. Mikami, T. Hosokawa, and T. Hasegawa, "Advantage of dichromatis over trichromats in discrimination of color-camouflaged stimuli in humans," *Perceptual and Motor Skills*, vol. 102, pp. 3–12, 2006.

[18] A. D. Melin, L. M. Fedigan, H. C. Young, and S. Kawamura, "Can color vision variation explain sex differences in invertebrate foraging by capuchin monkeys?," *Current Zoology*, vol. 56, no. 3, pp. 300–312, 2010.

[19] A. D. Melin, C. Hiramatsu, L. Fedigan, M. M. Schaffner, F. Aureli, and S. Kawamura, "Polymorphism and Adaptation of Primate Colour Vision," in *Evolutionary Biology: Mechanisms and Trends* (P. Pontarotti, ed.), ch. 13, pp. 225–241, Heidelberg, New York, Dordrecht, London: Springer, 2012.

[20] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 136, pp. 97–136, 1980.

[21] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, pp. 219–227, 1985.

[22] L. Itti and C. Koch, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[23] L. Itti and C. Koch, "Computational modelling of visual attention.," *Nature Reviews Neuroscience*, vol. 2, pp. 194–203, Mar. 2001.

[24] L. Zhang, M. H. Tong, T. K. Marks, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, pp. 32:1–20, 2008.

[25] N. Bruce and J. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, pp. 1–24, 2009.

[26] A. Oliva and P. G. Schyns, "Diagnostic colors mediate scene recognition.," *Cognitive psychology*, vol. 41, pp. 176–210, Sept. 2000.

[27] G. Naor-Raz, M. J. Tarr, and D. Kersten, "Is color an intrinsic property of object representation?," *Perception*, vol. 32, no. 6, pp. 667–680, 2003.

[28] D. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 1150–1157 vol.2, 1999.

[29] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, 1999.

[30] A. Derrington, J. Krauskopf, and P. Lennie, "Chromatic mechanisms in lateral geniculate nucleus of macaque.," *The Journal of Physiology*, vol. 357, no. 1, pp. 241–265, 1984.

[31] M. Yoshida, L. Itti, D. J. Berg, T. Ikeda, R. Kato, K. Takaura, B. J. White, D. P. Munoz, and T. Isa, "Residual attention guidance in blindsight monkeys watching complex natural scenes.," *Current Biology*, vol. 22, pp. 1429–34, Aug. 2012.

[32] J. Neitz, M. Neitz, J. C. He, and S. K. Shevell, "Trichromatic color vision with only two spectrally distinct photopigments," *Nature Neuroscience*, vol. 2, pp. 884–8, Oct. 1999.

[33] T. Wachtler, U. Dohrmann, and R. Hertel, "Modeling color percepts of dichromats," *Vision Research*, vol. 44, pp. 2843–55, Nov. 2004.

[34] P. DeMarco, J. Pokorny, and V. C. Smith, "Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats.," *Journal of the Optical Society of America. A*, vol. 9, pp. 1465–76, Sept. 1992.

[35] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman, "The CIECAM02 Color Appearance Model," in *Color Imaging Conference*, pp. 23–27, 2002.

[36] W. J. Ma, V. Navalpakkam, J. M. Beck, R. V. D. Berg, and A. Pouget, "Behavior and neural basis of near-optimal visual search.," *Nature Neuroscience*, vol. 14, pp. 783–90, June 2011.

[37] C. A. Párraga, T. Troscianko, and D. J. Tolhurst, "The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multiresolution model.," *Vision Research*, vol. 45, pp. 3145–3168, Nov. 2005.

[38] S. Tajima and M. Okada, "Discriminating natural image statistics from neuronal population codes," *PLOS ONE*, vol. 5, p. e9704, Jan. 2010.

[39] R. D. Valois, D. D. Albrecht, L. L. Thorell, and R. De Valois, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, vol. 22, no. 5, pp. 545–559, 1982.

[40] D. D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Physical review letters*, vol. 73, no. 6, pp. 814–817, 1994.

[41] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, pp. 2379–94, Dec. 1987.

[42] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience*, vol. 9, pp. 181–107, 1992.

[43] D. MacKay, *Information theory, inference and learning algorithms*. 2003.

[44] S. Ishihara, *Tests for colour-blindness*. 1917.

[45] J. Perez-Carpinell, R. Baldovi, M. D. de Fez, and J. Casro, "Color memory matching: Time effect and other factors," *COLOR Research and Application*, vol. 23, no. 4, pp. 234–247, 1998.

[46] D. H. Brainard, "The Psychophysics Toolbox.," *Spatial Vision*, vol. 10, pp. 433–436, Jan. 1997.

[47] D. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spatial Vision*, vol. 10, no. 4, pp. 437–442, 1997.

[48] N. G. Caine and N. I. Mundy, "Demonstration of a foraging advantage for trichromatic marmosets (Callithrix geoffroyi) dependent on food colour," *Proceedings of The Royal Society of London B*, vol. 267, no. 1442, pp. 439–444, 2000.

[49] A. Melin, L. Fedigan, C. Hiramatsu, C. Sendall, and S. Kawamura, "Effects of colour vision phenotype on insect capture by a free-ranging population of white-faced capuchins, Cebus capucinus," *Animal Behaviour*, vol. 73, pp. 205–214, Jan. 2007.

[50] A. C. Smith, A. K. Surridge, M. J. Prescott, D. Osorio, N. I. Mundy, and H. M. Buchanan-Smith, "Effect of colour vision status on insect prey capture efficiency of captive and wild tamarins (Saguinus spp.)," *Animal Behaviour*, vol. 83, pp. 479–486, Feb. 2012.

[51] M. S. Caywood, B. Willmore, and D. J. Tolhurst, "Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning.," *Journal of neurophysiology*, vol. 91, pp. 2859–73, June 2004.

[52] S. Tajima, H. Takemura, I. Murakami, and M. Okada, "Neuronal population decoding explains the change in signal detection sensitivity caused by task-irrelevant perceptual bias," *Neural Computation*, vol. 22, pp. 2586–2614, Oct. 2010.

[53] M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation.," *Nature reviews. Neuroscience*, vol. 13, pp. 51–62, Jan. 2012.

[54] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Enhancing by saliency-guided decolorization," in *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 257–264, Ieee, June 2011.

Satohiro Tajima received the B.A. and M.S. degrees from the University of Tokyo, Japan, in 2007 and 2009, respectively. During working at Japan Broadcasting Corporation (Nagano Station and Science Technology Research Laboratories), he received the Ph.D. degree in engineering from the University of Tokyo in 2013. He is studying theoretical neuroscience and cognitive science as a postdoctoral fellow, visiting researcher and visiting scientist at Brain Science Institute, RIKEN, Japan from 2013, and as a postdoctoral fellow at Department of Neuroscience, University of Geneva, Switzerland from 2014. He is a recipient of JSPS postdoctoral fellowship from 2014.

Kazuteru Komine received his B.E. and M.E. degrees from Tohoku University, Miyagi, Japan and his Ph.D. from Tokyo Institute of Technology, Tokyo, Japan in 1990, 1992 and 2008 respectively. He joined NHK (Japan Broadcasting Corporation) in 1992, and successively worked for NHK Science and Technical Research Laboratories since 1994. His major research interests include cognitive science of human vision, human factors of information displays and user interfaces for information systems. He is currently a senior manager of Three-Dimensional Image Research Division of STRL. He is a member of ITE (the Institute of Image Information and Television Engineers of Japan) and IEICE (the Institute of Electronics, Information and Communication Engineers).