Improving Calibration Time and Accuracy for Situation-Specific Models of Color Differentiation

David R. Flatla and Carl Gutwin
Department of Computer Science, University of Saskatchewan
110 Science Place, Saskatoon, Canada, S7N 5C9
david.flatla@usask.ca, gutwin@cs.usask.ca

ABSTRACT

Color vision deficiencies (CVDs) cause problems in situations where people need to differentiate the colors used in digital displays. Recoloring tools exist to reduce the problem, but these tools need a model of the user's color-differentiation ability in order to work. Situation-specific models are a recent approach that accounts for all of the factors affecting a person's CVD (including genetic, acquired, and environmental causes) by using calibration data to form the model. This approach works well, but requires repeated calibration – and the best available calibration procedure takes more than 30 minutes. To address this limitation, we have developed a new situation-specific model of human color differentiation (called ICD-2) that needs far fewer calibration trials. The new model uses a color space that better matches human color vision compared to the RGB space of the old model, and can therefore extract more meaning from each calibration test. In an empirical comparison, we found that ICD-2 is 24 times faster than the old approach, and had small but significant gains in accuracy. The efficiency of ICD-2 makes it feasible for situation-specific models of individual color differentiation to be used in the real world.

Categories and Subject Descriptors: K.4.2 [**Social Issues**]: Assistive technologies

General Terms: Human Factors

Keywords: Color vision deficiency (CVD), color blindness, color differentiation, adaptation tools, modeling

1. INTRODUCTION

Approximately ten percent of people [23] have some form of color vision deficiency (CVD - commonly called color blindness). People with CVD experience *color confusion*, in which they cannot distinguish between colors that are distinct for individuals without CVD. Color confusion leads to difficulties understanding information that is encoded using color. These difficulties range from annoyances (e.g., not being able to distinguish visited from unvisited hyperlinks), to critical safety issues (e.g., distinguishing traffic signs and signals, recognizing warning messages).

In the digital domain, color confusion problems can be at least partially addressed through adaptation tools that recolor images. These tools modify images to use colors that are more differentiable for individuals with CVD; in order to do this, they rely on models of color differentiation, both to identify problem colors and to find replacement colors.

ASSETS'11, October 24–26, 2011, Dundee, Scotland, UK.

Traditional differentiation models rely on many assumptions about the user and the environment that greatly limit their applicability. For example, these models assume that the user has a particular type of color vision deficiency called dichromatism, and that the user is working in an environment with constant and controlled lighting – assumptions that do not hold for other types of CVD or other environmental situations.

A new approach that solves the problem of assumption-based models is called *situation-specific modeling (SSM)* [8]. This technique builds the color-differentiation model not from assumptions, but from empirical data gathered from the user in the actual environment where they will use the model. SSM uses performance or judgment tasks to determine exactly what the user can and cannot differentiate in the current setting, thus implicitly accounting for all factors that affect the user's color vision. SSM is far more sensitive than assumption-based models [8]; however, the main limitation of the approach is that the accuracy of the model degrades as conditions change (e.g., as the user becomes tired, or lighting changes in the room), and eventually another calibration must be carried out. The best current SSM technique (called ICD [8]), however, requires more than thirty minutes for calibration, limiting the applicability of the approach.

To address this problem, we have developed a new SSM model (ICD-2) that requires a much shorter calibration. The new model is a substantial improvement because it is based on a much more complete understanding of the mechanics of color vision than the previous model's RGB color space. The old model measured minimum differentiability on each of the red, green, and blue channels, and working solely in RGB space means that many such measures must be taken because RGB is not *perceptually uniform* (i.e., the RGB difference between any two colors does not correlate with their visually-perceived difference). ICD-2, in contrast, represents color differences using the LUV color space (which is perceptually uniform), allowing the measurements for a single color to be generalized to all other colors. In addition, ICD-2 measures color differentiation abilities in terms of *color confusion axes*, which embody the exact color confusion difficulties experienced by individuals with CVD [28]. These two underlying changes to the approach mean that ICD-2 requires only eight color differentiation measurements, compared with the 192 measurements needed by the previous model.

To validate ICD-2, we compared it to the old ICD model in a study with both CVD and non-CVD participants. The study calibrated and built both the old and new models for each person, and then tested the color-differentiability predictions generated by the models against a ground truth set established by the participant. We found that calibration with ICD-2 was 24 times faster than with the old model (2.17 minutes vs. 52.6 minutes), and that the ICD-2 model was modestly but statistically significantly more accurate (78.7% vs. 76.1%).

We make three main contributions in this paper. First, we demonstrate that the situation-specific modeling approach can be feasible and practical if it has an efficient model such as ICD-2. Second, we show how calibration time and model accuracy can be traded off, which provides an additional degree of flexibility for SSM. Third, we improve the understanding of how different types of color-differentiation models affect the larger problems of recoloring and reducing color confusion for people with CVD.

2. BACKGROUND

2.1 Color Vision Deficiencies

There are several sources of color vision problems, ranging from genetic disorders to acquired deficits to situation-specific environmental factors.

- *Genetic Factors* can cause anomalies or deficiencies in the three types of color-sensing cells of the retina (cones). Different types of cones are sensitive to different parts of the visible spectrum: *protan* cones for long-wavelength, *deutan* cones for medium-wavelength, and *tritan* cones for short-wavelength light. *Anomalous trichromacy* can result if these wavelength sensitivities are shifted for some cones. If all cones of a certain type are missing, *dichromacy* can result. In rare cases, people can be missing two types of cones (*cone monochromacy*) or all color-sensing cells (*rod monochromacy*), limiting perception to shades of grey [1][5].
- *Acquired Deficits* involve damage to the vision system from external events such as accident, disease, or exposure to harmful chemicals. Retinopathy occurs when a portion or all of the photoreceptors of the retina die – e.g., resulting from diabetes or long-term exposure to styrene. When photoreceptors die, color perception can be drastically altered [14]. Aging can also bring changes to color vision: for example, yellowing of the lens and cataracts both modify the light entering the eye.
- *Situational Effects* are factors in the user's local environment that cause temporary changes to color perception; if the resulting effect causes problems, these are known as 'situationally-induced disabilities' [22]. These factors can be of many different types, and can include characteristics of the color source (e.g., the quality and calibration of the monitor that displays the colors), physical characteristics of the environment (e.g., the amount of ambient light in the room, or glare from light striking the display), or the state of the user (e.g., the presence of drugs such as Viagra or antidepressants).

All types of CVD cause similar problems for our purposes – they make it difficult for people to differentiate among colors that can be distinguished in other circumstances or by other users. The use of color in digital information displays is ubiquitous, and many presentations require that the user be able to tell different colors apart. For example, information visualization uses color for categorical encoding, highlighting, popout, and representation of continuous variables [24][27]. Everyday interfaces also use color extensively – for example, to show visited links in web pages, or to clearly indicate alerts and warnings.

2.2 Color-Adaptation Systems and Models

Several systems exist to aid the problem of color differentiation in digital displays. These techniques select colors in an image that are likely to be problematic for the user, and switch these to different colors that are more likely to be differentiable. From an early proposal by Meyer and Greenberg [16], several methods have been developed including SmartColor [26] and a number of techniques for dealing specifically with photographic images (e.g., [12][19][20]). In some cases the user can participate in the recoloring process – for example, some forms of CVD (such as anomalous trichromatism) are less severe than dichromatism, and an interactive system can allow these users to guide the recoloring process [10][11].

Regardless of their approach, all recoloring techniques rely on an underlying model of the user's color differentiation abilities, and most current models are based on an early assumption-based algorithm [4][16][25]. This algorithm allows the simulation of dichromatic color perception for individuals without CVD: it first transforms an image from RGB to the Long-Medium-Short (LMS) cone stimulation color space, then removes appropriate wavelength information for the desired type of dichromatism (e.g., long-wavelength for protanopes), then translates the modified LMS colors back to RGB. The algorithm can be used to detect color confusion by comparing colors in the original and modified images – if regions that were different colors become the same, the colors will not be differentiable.

This algorithm uses several assumptions, however, that limit the applicability of the model. The transformation to LMS requires that the emission spectra for the monitor are known (which varies widely across monitor technologies), and that the monitor is calibrated in terms of white balance and gamma. The transformation also assumes a 'representative' human color vision system, but individual differences are common [17]. Most importantly, the model does not deal with the variability seen in CVD users (e.g., in anomalous trichromats), nor does it handle other forms of CVD such as extreme anomalous trichromacy [3] or monochromacy. It also does not take into account any of the acquired or situational causes of CVD as described above. There is at least one model proposed that handles anomalous trichromacy [15], but it requires details about how far the peak wavelength of the photoreceptor of interest has been shifted, which is not easily obtained.

To address these limitations, Flatla and Gutwin [8] developed *situation-specific models* of color differentiation. SSMs account for all of the factors affecting a user's color perception, by testing their differentiation abilities in the actual scenario and environment of use, with performance or judgment tests. More details about this approach will now be presented.

3. COLOR DIFFERENTIATION MODELS

Color differentiation models provide information about the differentiability of colors. This can be accomplished through a simple API that provides a single function:

```
boolean : areDifferentiable(Color c1, Color c2)
```

This function accepts two colors and returns true if the model predicts that they are differentiable and false otherwise. To accomplish this, the function predicts the set of colors that are not differentiable from c1, and then checks to see whether c2 is in this set. If it is not in the set, then the colors are differentiable.

Certain color spaces (e.g., RGB, CIE XYZ, CIE LUV) all exhibit a property that is beneficial for identification of these 'not differentiable' sets. Considering color spaces as physical spaces (in which each color occupies a unique location) allows the metaphor of movement through a color space. As one moves away from any particular color, increasingly different colors are encountered. While moving on any path away from a color, a point will eventually be reached at which the colors that are being encountered transition from not differentiable to differentiable, in reference to the starting color. This transition is called a psychometric function, which can be described mathematically using a sigmoid function. For simplicity, many models (including ICD and ICD-2) approximate the psychometric function as a step function, which transitions from 'not differentiable' to 'differentiable' in a single step (Figure 1). When reasonable paths are chosen, these transition points (called *differentiation limits*) can be used to define a volume around the starting color. All colors within this volume are predicted to be not differentiable from the starting color, and all colors outside the volume are predicted to be differentiable.

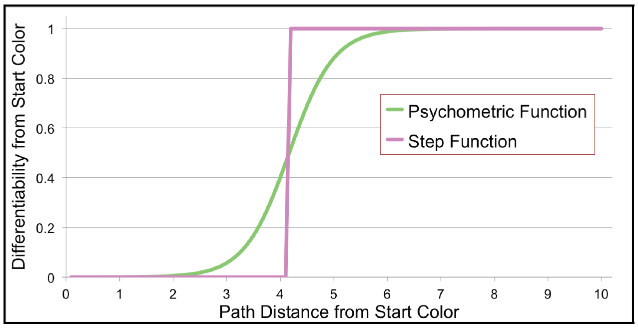

Figure 1. Approximation of a normalized psychometric sigmoid function with a simple step function.

3.1 Original ICD Model (the 'old model')

The situation-specific model proposed in [8] (ICD, here referred to as the 'old model') relies exclusively on the RGB color space, in which each channel (red, green, and blue) defines an orthogonal axis in a 3D space. For any particular color, six lines extend out from the color, two for each channel axis (one in the increasing direction, the other in the decreasing direction). Along each of these lines, a differentiation limit can be identified where the colors along the line become differentiable from the original color. These six limits (one for each line) are used to define a box around the original color. Colors within this box are considered not differentiable from the original color, and colors outside are considered differentiable.

3.1.1 How to calibrate the old model?

As RGB is not a perceptually uniform color space, the differentiation limits for one color do not generalize well to other colors. To calibrate the old model, therefore, differentiation limits are measured for many colors (64) uniformly spread through the RGB color cube. To measure a single differentiation limit, binary search is used along the color channel line described above. The user is presented with a rectangular field of dots on a black background (Figure 2). One half of the dots are the starting color (the color for which differentiation limits are needed). The other half of the dots are a color that lies along the line between the original color and the extreme value for the channel involved. The user responds with either 'not different' or 'different' depending on whether he/she sees a difference between the two colors. If the user says 'not different' then the colors are made more different and redisplayed. If the user says 'different' the colors are made less different and redisplayed. This is repeated in a binary search pattern, until the differentiation limit is identified.

This binary search process requires approximately ten presentations (taking about one second each) before a single differentiation limit is identified. For reasons presented in [8], only increasing differentiation limits are required for calibration, giving three limits per calibration color (of which there are 64). As a result, calibrating the old model requires 64x3x10=1920 samples – if each takes one second, a total of 32 minutes.

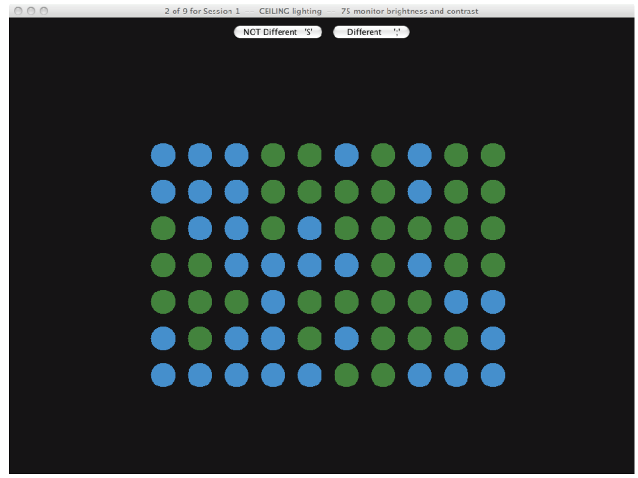

Figure 2. Screen presented during original calibration. User responds with 'not different or 'different' [8].

3.1.2 How to make predictions?

To answer an 'areDifferentiable(Color c1, Color c2)' query, the model uses trilinear interpolation to estimate the differentiation limits for one of the parameters. The other color parameter is then tested to see whether it is inside or outside of the box defined by these interpolated limits.

3.2 ICD-2 Model

The old model requires a lengthy calibration because differentiation limits are required for 64 RGB colors. As identified above, this is because RGB is not perceptually uniform. By using a perceptually uniform color space (CIE LUV [6]), the number of limits necessary for calibration can be drastically reduced.

Images are encoded using the RGB space in digital environments, so moving colors in RGB space to LUV space is necessary for using the LUV color space in the new model. To accomplish this, we use the sRGB transform provided by [13] to move between RGB and CIE XYZ, and transforms from the same source to move between XYZ and LUV. As a result, transformations from RGB to LUV (and back again) can be accomplished.

CIE LUV is a color space that is perceptually uniform and separates the description of a color into a luminance axis (L – ranging from 0-100) and two chromaticity coordinates (U,V – centered at 0). This allows colors of equal luminance to be found simply by holding L constant and varying U and V. When U and V equal 0, the color is achromatic (black, grey, or white).

In the LUV color space, the differentiation abilities of an individual can be described well using a *discrimination ellipse* when only colors of identical luminance are considered [21]. For a given color, a discrimination ellipse can be found that surrounds the color. Those colors outside the ellipse are differentiable and those inside are not differentiable from the original color. To extend this to a three-dimensional shape, *discrimination ellipsoids* [18] are used in the ICD-2 model. The discrimination ellipsoid is defined using the discrimination ellipse and two points above and below the equal luminance plane. The ellipsoid that matches the ellipse and intersects the two points is the ellipsoid used.

To find the discrimination ellipse, six discrimination limits are measured via a calibration procedure. Instead of finding these limits along RGB channel axes, as in the old model, we find these limits along three *lines of confusion* [28]. A line of confusion is defined by a base color and a *copunctal point*. The colors that lie along each line are not differentiable for individuals with dichromatic CVD (Figure 3) and there is one copunctal point for each type of dichromacy. Each line gives two differentiation limits, one moving from the base color to the copunctal point, and one moving from the base color away from the copunctal point, resulting in six differentiation limits. These six limits are then used to generate the best-fit ellipse using approaches outlined in [7] and [9]. The half lengths of the major and minor axes of this best-fit ellipse are then used to find the ellipsoid.

Figure 3. Confusion line colors for individuals with tritanopia (blue-yellow dichromatic CVD) using the base color for the ICD-2 model: a) confusion colors toward the tritan copunctal point, b) tritan simulation of (a), c) confusion colors away from the tritan copunctal point, d) tritan simulation of (c).

To find the two points above and below the luminance plane, two additional discrimination limits are measured via the calibration procedure. These limits correspond to the amount of luminance that needs to be added and subtracted from the base color in order for the user to perceive a difference.

An ellipsoid can be described using the following formula:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$$

where a is the best-fit ellipse major axis half length, b is the best-fit ellipse minor axis half length, and c is the amount of luminance that is added or subtracted, as described above. This formula is used in ICD-2 to internally represent the discrimination ellipsoid.

3.2.1 How to calibrate ICD-2?

To calibrate the ICD-2 model, eight discrimination limits are needed for a single base color. These limits are found in a similar manner as the old model, but the technique has been modified so that the user no longer provides a judgment about the differentiability of the two colors presented, but rather performs a task. If the user can perform the task, then it is interpreted that the user can see the difference between the two colors. If they cannot do the task, then it is interpreted as the user not being able to differentiate between the two colors. The base color was chosen to be the color represented by the LUV coordinates L=50.0, u=0.0, v=0.0, which is a mid-luminance grey. These LUV coordinates map to RGB color (118,118,118) using the sRGB transform described above.

This performance task involves the user identifying the orientation of a circle with 1/8th of its perimeter missing (Figure 4). The circle is presented to the user, and if they can identify the location of the gap, they press a correspondingly labeled key on the numeric keypad. If they see no gap, the user presses the space bar. A binary search approach is used to find the differentiation limit, where the difference between the circle and the background is increased when the user sees no difference, and decreased when they see a difference.

To facilitate the performance task, the presentation of the colors to the user has been modified to approximate the approach used in [21] to determine discrimination ellipses. A 400x400 pixel region on a black screen is presented to the user. This region is filled with a regular pattern of small (4-pixel diameter) circles, with black between the circles. The gapped circle introduced above is superimposed on this background of small circles, such that the background is the base color, and the gapped circle is a color along the confusion line (or luminance line) for which a discrimination limit is desired (Figure 4). The numeric keypad of the keyboard was modified with labels such that the labels matched the possible orientations of the gapped circle.

When two colors are placed directly adjacent to each other, any differences in luminance between the two colors results in the user seeing a difference between the two colors, even though this difference may go away as soon as a small gap is introduced between the two colors. To offset the effect of luminance contrast, temporal random luminance noise [1] was applied to the entire presentation (background and gapped circle). This noise produces colors with identical LUV UV coordinates (chromaticity), but varying L values (luminance). The black space between the small circles further reduces the effects of luminance contrast.

When a differentiation limit has been identified, its Euclidean distance in LUV space is recorded. This gives a distance along each confusion line (6 measures) and distance in luminance above and below the base color (2 measures).

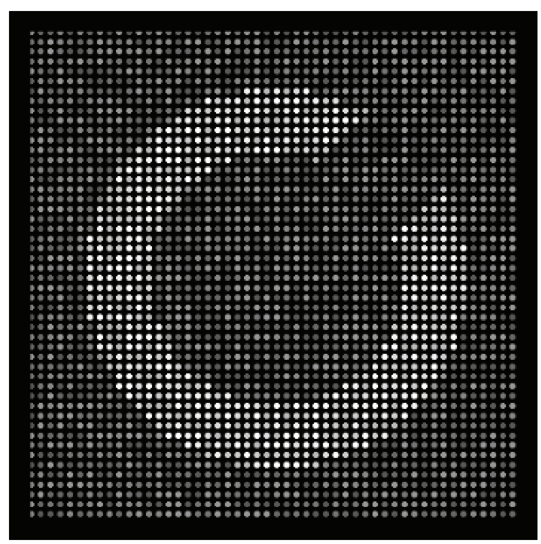

Figure 4. Screen presented during ICD-2 model calibration. User presses numeric keypad key that matches the rotational orientation of the gapped circle, or space if they see nothing.

3.2.2 How to make predictions?

To answer an 'areDifferentiable(Color c1, Color c2)' query, the model first converts both colors to LUV color space (using the default sRGB transform mentioned above), and determines which is closer to the base color using Euclidean distance. The closest color is chosen because the discrimination limits for the base color are known, and if the LUV color space is not perceptually uniform for those with CVD, then the discrimination limits for the color that is closest to the base color should be more similar to the base color differentiation limits. The parameter nearest to the base color is called the primary color, the other parameter is called the secondary color. To determine an appropriate discrimination ellipsoid, the differentiation ellipsoid for the base color needs to be transformed to fit the primary color.

To save computation, before the discrimination ellipsoid is transformed, a luminance comparison is performed. Using the luminance thresholds determined during the calibration for the base color, luminance bounds for the primary color are determined. If the luminance of the secondary color is outside of these bounds, the query returns ‘differentiable’. If the luminance of the secondary color falls within these bounds, then the transformation is performed.

To transform the discrimination ellipsoid of the base color to the primary color, the confusion lines for the primary color are found (between the primary color and the copunctal point). As confusion lines are defined by the copunctal point and another point, they are not rotationally invariant as different ‘other’ points are selected. Once the color confusion lines for the primary color are found, the algorithm walks along each confusion line away from the primary color until the LUV distance just exceeds the differentiation limit LUV distance found during the calibration. The LUV coordinates for the color at this point on the confusion line are then used to specify six points in the UV plane for the primary color. These six points are used to find the best-fit ellipse.

Using the luminance of the secondary color, the best-fit ellipse is resized to be the discrimination ellipse for the primary color, but at the secondary color’s luminance, by using modifications of the formula for an ellipsoid given above to find the adjusted half major and half minor axis lengths for the resized ellipse:

$$a' = \sqrt{a_p^2 - \frac{a_p^2 * (L_s - L_p)^2}{c_p^2}} \qquad b' = \sqrt{b_p^2 - \frac{b_p^2 * (L_s - L_p)^2}{c_p^2}}$$

Figure 5. Formula for the resized half major axis length (left) and the formula for the resized half minor axis length (right).

Once the resized ellipse is found, its center, dimensions, and orientation are used to transform a unit Java `Ellipse2D.Double` object. When transformed, this object is defined by a `Path2D.Double` object, which provides a `contains(Point2D.Double)` method, which determines if the given point is within the ellipse. The LUV UV coordinates for the secondary color are packaged into a `Point2D.Double` object and passed as a parameter to the ‘contains’ method. If the point is in the ellipse, then the query returns ‘not differentiable’, otherwise it returns ‘differentiable’.

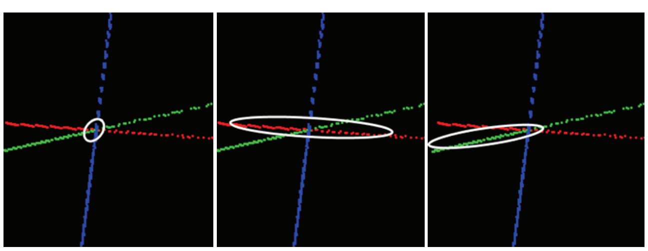

Figure 6. Three examples of the ellipses found by ICD-2: normal color vision (left), protan CVD (middle), deutan CVD (right). Colored lines indicate confusion axes.

4. EVALUATION

We compared the ICD-2 model to the original ICD model [8] (here called the ‘old model’) in an empirical study. There were two main goals of our evaluation – first, to confirm that calibration of ICD-2 is in fact faster than the old model (and to determine the actual reduction in time), and second, to determine ICD-2’s accuracy compared with the existing approach.

4.1 Study Methods

To compare the ICD-2 model with the old model, we conducted a user study with 16 male participants (mean age 33.8 years) - eight who self-identified as having CVD (mean 39.0 years) and eight who self-identified as not having CVD (mean 28.7 years). As both the old model and ICD-2 are general models of color differentiation, we did not perform any tests to assess the type or severity of participant CVD. We constructed a custom Java application using the Processing libraries for displaying visual content to the screen (processing.org). The study ran in a single location on a Windows 7 machine using a 20-inch 1600x1200 Dell 2001fp monitor.

During the study, participants performed two tasks. The first task collected calibration data for generating both models. In the second task, ‘ground truth’ responses were collected from the participant to evaluate each generated model.

4.1.1 Calibration Task

As the study was designed to compare the models to each other, identical calibration procedures were used for the old model and the ICD-2 model. We opted to use the calibration procedure with the gapped circle (described above) for both models. The procedure was modified for the old model in order to gather increasing and decreasing differentiation limits on RGB channels.

As the old model calibration is time consuming, a reduced set of old model calibration points was gathered to reduce study run time. Nine points were chosen in RGB space to approximate the uniform spread of the 64 calibration points in the true calibration of the old model. These were (118,118,118), and eight additional colors, one halfway along each ray from this start color to the eight corners of the RGB color cube. This gave the following nine colors: grey, black, green, yellow, red, blue, purple, cyan, and white. For each of these, six differentiation limits were collected from the user, for a total of 54 differentiation limits. To calibrate the ICD-2 model, the standard eight differentiation limits around (118,118,118) were collected. This gave a total of 62 differentiation limits.

The order of these limits was randomized and presented to the participant sequentially. When the participant supplied a response (either the space bar for ‘no circle visible,’ or the appropriate numeric keypad key) the difference between the background and the circle was adjusted accordingly and the limit was reinserted into the sequence. If an incorrect numeric key was pressed, it was interpreted the same as pressing the space bar. Once the participant had given a response for each of the 62 limits, the order was shuffled and presented sequentially to the participant again. This was repeated until the binary search for each limit converged on a single value. For the old model, this value was reported as a raw RGB channel difference. For the ICD-2 model, this value was converted into its equivalent LUV Euclidean distance from the base color. The entire calibration required about 400 presentations to the participant, taking approximately 30 minutes. The participant could take a break at any time, but was encouraged to take at most 3-4 seconds per presentation.

The total time to collect each differentiation limit was recorded as well. With this time data, the total time to gather the 54 old model differentiation limits, and the total time to gather the eight ICD-2 calibration differentiation limits was measured. As the original old model requires 192 calibration limits, the actual time to collect the 54 limits was scaled up (192 * time / 54) to reflect the original calibration time.

4.1.2 Evaluation Task

Once the calibration data was collected, the participant took a short (~5 minute) break to rest their eyes. Once finished, the experimenter conducted the evaluation test. At the beginning of this task, the calibration data from the first session was used to generate the old model and the ICD-2 model. These models were then used to generate evaluation trials as described below.

Any two RGB colors selected randomly have a high probability of being differentiable. We wanted to use evaluation data that would provide a more uniform chance of each model predicting that the colors would be differentiable or not. To accomplish this, the models based on the calibration data for each participant were used to generate the evaluation trials for that participant. Using the nine RGB colors mentioned above (grey, black, green, yellow, red, blue, purple, cyan, white), each model was asked to generate two sets of 15 colors – one that the model predicts as being differentiable from the supplied color and one that the model predicts as being not differentiable from the supplied color. To accomplish this, colors were uniformly randomly selected from a volume twice as large as the old model box or ICD-2 ellipsoid. Each color selected was then predicted as differentiable or not differentiable and added to the appropriate set. The sets were returned when they were both full. This resulted in 15x2x2x9=540 trials for the evaluation session. These trials were randomly presented to the participant using the same test procedure (with the gapped circle) as the ICD-2 calibration. If the participant correctly identified the orientation of the gapped circle, then the colors were recorded as 'differentiable'; otherwise 'not differentiable' was recorded. This was used to establish a 'ground truth' set of color comparisons. For each of these responses, each models' differentiability prediction was also recorded.

4.2 Study Design

Our evaluation used a repeated-measures factorial design with two factors: *model type* (old model or ICD-2) and *CVD presence* (normal color vision or CVD). The CVD-presence factor was used only to check for interactions in the other analyses.

Four dependent variables were recorded by the system. *Calibration time* was gathered from the calibration task as described above. The remaining three variables were calculated from the raw correct / incorrect data gathered from comparing the predictions to ground truth: *overall accuracy* (number of correct predictions over total trials), *false positive rate* (proportion of predictions that incorrectly suggested that colors were differentiable), and *false negative rate* (proportion of predictions that incorrectly suggested that colors were not differentiable).

Our analysis used repeated-measures ANOVAs to test the effects of model type on these four dependent variables, and to look for interactions with CVD presence.

4.3 Results

4.3.1 Calibration Time

We recorded the time needed to carry out the entire calibration with both the old and new models, and scaled the old model calibration time to reflect the true calibration procedure for the old model (from 54 to 192 limits). As shown in Figure 7, calibration time for ICD-2 (mean of 2.17 minutes) is dramatically lower than for the old model (mean 52.6 minutes). Not surprisingly, the effect of model type is significant ($F_{1,14}$=123.46, $p<0.001$); there was no interaction with CVD presence ($F_{1,14}$=0.077, p=0.78). The 24-times improvement is proportional to the reduction in the number of calibration trials (from 192 to 8). It should be noted that these times are for the 'gapped circle' calibration technique, which is slower than the old calibration technique (see Figure 2); with the old technique, it is likely that the calibration time for ICD-2 will be even lower.

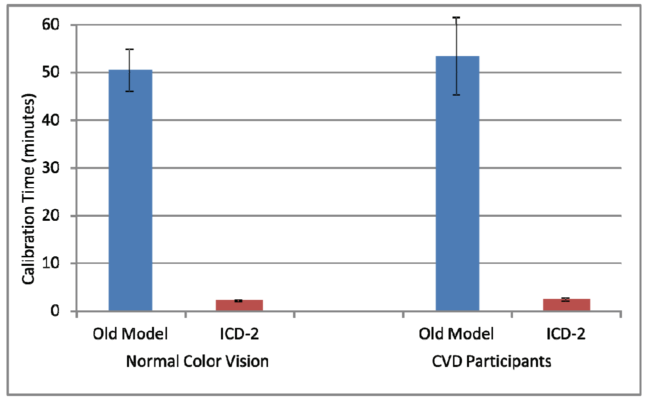

Figure 7. Calibration time ±s.e, by CVD status and model.

4.3.2 Model accuracy

As described above, we tested the accuracy of the models' predictions by comparing them to the ground truth of the 540 evaluation trials collected from participants.

Overall Accuracy. The overall mean accuracy for the old model was 76.1%, and for ICD-2 was 78.7% (see Figure 8). ANOVA showed that model type had a significant main effect on accuracy ($F_{1,14}$=5.13, $p<0.05$), with ICD-2 at approximately 2.6% higher accuracy. There was no interaction with CVD presence ($F_{1,14}$=1.15, p=0.30).

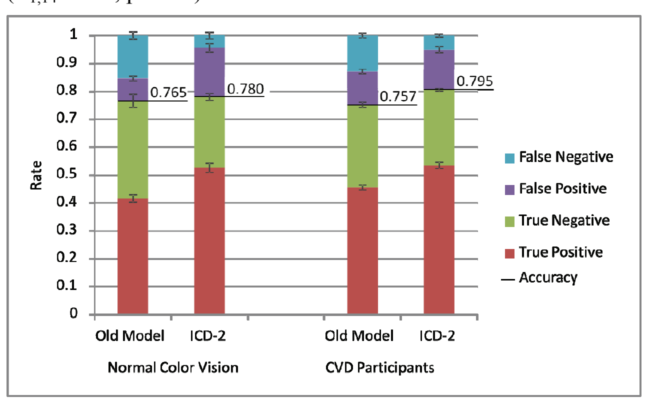

Figure 8. Model accuracy ±s.e, by CVD status and model (true positive rate + true negative rate = overall accuracy).

False Positive Rate. The mean false positive rate for the old model was 10.01%, and for ICD-2 was 16.02%. ANOVA shows that this difference is significant ($F_{1,14}$=64.19, p<0.001); ICD-2 had approximately 6% more false-positive errors. In addition, there was a significant interaction between model type and CVD presence ($F_{1,14}$=22.16, p<0.001). As shown in Figure 8, the false-positive difference between the old and new models is larger for participants with normal color vision than for CVD participants. We discuss the implications of these differences below.

False Negative Rate. Mean false negative rates were 14.18% for the old model and 4.75% for ICD-2. This difference is also significant ($F_{1,14}$=154.17, p<0.001). There was no interaction with CVD presence ($F_{1,14}$=4.23, p=0.059).

5. DISCUSSION

Our evaluation of ICD-2 provided four main results:

1. Calibration of ICD-2 is dramatically faster than the old model, requiring 1/24 the time of the old calibration;
2. ICD-2 is significantly more accurate than the old model, with a 2.6% improvement in overall accuracy;
3. ICD-2 does show a higher rate of false positives (although primarily for participants with normal color vision).

In the next sections we provide explanations for these results, discuss their implications for the feasibility of recoloring tools based on ICD-2, and outline ways in which the model could be improved and extended.

5.1 Explanation of main results

Calibration Time. The reason for the reduced time to calibrate ICD-2 is simple – using the perceptually uniform LUV color space and basing the discrimination volume on known color confusion lines means that ICD-2 requires far fewer calibration samples than the old model. These changes show the value of building the model on principles that more completely characterize human color perception. The short calibration time of ICD-2 (just over two minutes) means that we can even consider taking additional samples to further improve accuracy (as discussed below).

Overall Accuracy. The increase in accuracy for ICD-2 compared with the old model was modest (2.5%), so it is difficult to conclusively determine the source of the improvement. However, we believe that the change from a bounding box (in the old model) to an ellipsoid (in ICD-2) is the main reason for the better performance: previous research has shown that an ellipsoid better matches the way that humans perceive color and the way that individuals with CVD have difficulty with differentiation [18].

Higher False Positive Rate for Normal Color Vision. The ICD-2 model showed a 6% higher false-positive rate than the old model, for participants with normal color vision. False positive errors are more serious for recoloring tools than false negatives, since false positives lead to situations where the recoloring algorithm is not able to identify problem colors (because two colors that are actually not differentiable are predicted to be). The seriousness of these situations is compounded further if the recoloring tool proposes replacement colors that are not in fact differentiable by the user. The higher false positive rate is directly caused by the model ellipsoid being smaller than it should be; the reasons why the model chose too-small ellipsoids, however, are not clear. One possibility is that the step between colors along the color confusion lines is too large. As these colors were pre-computed (to save processing time), it is possible that the chosen step was too large. This could result in the calibration returning a differentiation limit that was on the 'not differentiable' side of a step, even though the true differentiation limit was somewhere in the middle of the step. This would result in unnecessarily small differentiation limits, leading to a small ellipsoid.

A step size that is too large would also explain the difference in false positives between CVD and non-CVD participants. The error introduced by this problem would have an additive (not multiplicative) effect on the volume of the resulting model. Smaller volumes (e.g., for non-CVD users) would be more greatly affected by reducing their axes by a fixed amount than larger volumes (e.g., for CVD users).

5.2 Generalization

Application to Recoloring Tools. The problem of false positives can be dealt with through an 'offset factor' that arbitrarily increases the size of the ellipsoid (this factor was also needed for the old model [8]). The size of this offset differs per user, but can be easily calculated when the model is built. We note that the overall accuracy of the situation-specific modeling approach is such that even a liberal offset value will not greatly reduce the number of colors available to a recoloring algorithm.

Generalizing to Other Situations and Users. ICD-2 is able to generate different ellipses for different types of users; for example, Figure 6 shows ellipses for a normal user, a user with protan CVD, and a user with deutan CVD. This variation in generated models (and the associated accuracy results) present a strong argument that this modeling approach is applicable to many individuals with a variety of color differentiation abilities. In the future, we plan to examine how ICD-2 generalizes to different environmental situations, as well as internal variations, such as those associated with aging. The evaluation did include two individuals with CVD who were older (63 and 67 years old, one diagnosed with cataracts), who both experienced gains in accuracy from the old model to ICD-2 (3% and 2%, respectively). These results suggest that the model will generalize well, at least with regard to internal variations such as age and illness.

5.3 Improving and Extending the Model

Two main approaches present themselves for improving ICD-2, both of which involve the collection of more calibration points.

Increasing the Number of Differentiation Limits. The three confusion lines introduced above give rise to six differentiation limits defining the discrimination ellipse. To improve the shape, location, and size of the ellipse, additional differentiation limits at different points can be collected. These would be along lines of a different rotational orientation. Each new line would introduce two additional discrimination limits, so the accuracy of the model can be balanced against calibration time.

Increasing the Sampling of Single Differentiation Limits. For the eight differentiation limits used to define the ellipse, we noted above that we use step functions to approximate the true psychometric sigmoid function (Figure 1). If repeated samples of differentiation limits were collected, then the nature of this sigmoid could be determined. This would give two main benefits. First, the true sigmoid function could be used to overcome the 'step size' issue presented as an explanation for the increased false positive rate. Setting the differentiation limit to be the point where the sigmoid function levels off (at 1.0 differentiability), allows a false-positive-reducing ellipsoid to be constructed. Second, knowing the sigmoid function allows the model to provide a confidence score with each prediction. The model would have 100% confidence for where the sigmoid is 0.0 (not differentiable) and 1.0 (differentiable). In the region where the sigmoid rises from 0.0 to 1.0, the confidence can be determined by:

$$confidence = 4 * (sigmoid - 0.5)^2$$

Extending the Approach to Other Color Problems. The situation-specific modeling approach also shows promise for other kinds of color problems experienced by users with CVD. Two problems in particular that can use similar techniques are color matching and the effects of simultaneous color. Color matching is the task of finding a color that has been identified in another part of a display (e.g., using a legend in a bar chart to look for a particular data category). Simultaneous color issues arise when colors (and color differentiability) are affected by the presence of surrounding colors (e.g., colors on a dark background look lighter than they do on a light background). Performance-based SSMs can model both of these tasks and can further help designers to choose colors that work well both for CVD users and those with normal color vision. These models would also allow more broadly-applicable recoloring tools by providing the necessary information to successfully recolor images that contain matching and simultaneous contrast color use.

6. CONCLUSION & FUTURE WORK

Situation-specific modeling is a calibration-based approach to building empirical models of a user's color-differentiation ability. This approach is able to account for all of the factors affecting a person's CVD, including genetic, acquired, and environmental causes. The main difficulty with current situation-specific techniques is that the calibration procedure takes a long time, which is a problem because calibration must be carried out whenever the user's situation changes. In this paper we described a new SSM technique called ICD-2 that makes several improvements on the current state of the art. ICD-2 is based on a perceptually uniform color space that better matches human color perception, and so requires far fewer calibration steps than the old model; in addition, the change in color space also allows us to use known CVD color confusion lines to select our calibration points. We compared ICD-2 to the existing model in a controlled experiment, and found that ICD-2 was both dramatically faster for calibration and significantly more accurate than the old model. The efficiency and performance of ICD-2 makes it now feasible for situation-specific models of individual color differentiation to be used for recoloring tools in the real world.

Our future work in this area will follow three main directions. First, we plan to build and deploy a recoloring tool based on ICD-2, and test its performance in real-world use. Second, we plan several additional improvements to the model, including the development of confidence scores as discussed above. Third, we will extend our performance-based approach to other dimensions of color perception, such as color recognition or recall.

7. ACKNOWLEDGMENTS

Our thanks to Hervé Bitteur for providing the library for finding the best-fit 2D ellipse for a set of points.

8. REFERENCES

[1] Birch, J., Barbur, J., Harlow, A. New Method Based on Random Luminance Masking for Measuring Isochromatic Zones Using High Resolution Colour Displays. *Opthal. Physiol. Opt.* 12(2), 1992, 133-136.

[2] Birch, J. *Diagnosis of Defective Colour Vision*. 2nd ed., 2001.

[3] Birch, J. Extreme Anomalous Trichromatism. In Mollon, Pokorny, and Knoblauch, eds., *Normal and Defective Colour Vision*. Oxford University Press, 2003, 364-369.

[4] Brettel, H., Viénot, F., Mollon, J. Computerized Simulation of Color Appearance for Dichromats. *J. Opt. Soc. Am. A*., 14(10), 1997, 2647-2655.

[5] Cole, B. The Handicap of Abnormal Colour Vision. *Clinical and Experimental Optometry*, 87(4-5), 2004, 258–275.

[6] Commission Internationale de l'Eclairage (CIE). *Colorimetry*, 2nd ed., CIE Publication 15.2, 1986.

[7] Fitzgibbon, A., Pilu, M., Fisher, R. Direct Least Squares Fitting of Ellipses. *Trans. PAMI*, 21(5), 1999, 476-480.

[8] Flatla, D. R., Gutwin, C. Individual Models of Color Differentiation to Improve Interpretability of Information Visualization. *Proc. CHI 2010*, 2563-2572.

[9] Halir, R., Flusser, J. Numerically Stable Direct Least Squares Fitting of Ellipses. *Proc. WSCG 1998*, 59-108.

[10] Jefferson, L., Harvey, R. Accommodating Color Blind Computer Users. *Proc. ASSETS 2006*, 40-47.

[11] Jefferson, L., Harvey, R. An Interface to Support Color Blind Computer Users. *Proc. CHI 2007*, 1535-1538.

[12] Kuhn, G., Oliveira, M., Fernandes, L. An Efficient Naturalness-Preserving Image-Recoloring Method for Dichromats. *IEEE Trans. Vis. and Comp. Graphics* 14(6), 2008, 1747-1754.

[13] Lindbloom, B. Website: www.brucelindbloom.com. Last accessed May 6th, 2011.

[14] Lomax, R., Ridgway P., Meldrum M. Does Occupational Exposure to Organic Solvents Affect Colour Discrimination? *Toxicological Reviews*, 23(2), 2004, 91-121.

[15] Machado, G., Oliveira, M., Fernandes, L. A Physiologically-based Model for Simulation of Color Vision Deficiency. *IEEE Trans. Vis. & Comp. Graph.* 15(6), 2009, 1291-1298.

[16] Meyer, G., Greenburg, D. Color-Defective Vision and Computer Graphics Displays. *IEEE Comp. Graph. & Appl.,* 8(5), 1988, 28-40.

[17] Neitz J., Jacobs, G. Polymorphism of the Long-Wavelength Cone in Normal Human Colour Vision. *Nature*, 323, 1986, 623-625.

[18] Poirson, A., Wandell, B. The Ellipsoidal Representation of Spectral Sensitivity. *Vision Research*, 30(4), 1990, 647-652.

[19] Rasche, K., Geist R., Westall, J. Detail Preserving Reproduction of Color Images for Monochromats and Dichromats. *IEEE Comp. Graph. & Appl.,* 25(3), 2005, 22-30.

[20] Rasche, K., Geist R., Westall, J. Re-coloring Images for Gamuts of Lower Dimension. *Computer Graphics Forum*, 24(3), 2005, 423–432.

[21] Regan, B., Reffin, J., Mollon, J. Luminance Noise and the Rapid Determination of Discrimination Ellipses in Colour Deficiency. *Vision Research*, 34(10), 1994, 1279-1299.

[22] Sears, A., Lin, M., Jacko, J., Xiao, Y. When Computers Fade: Pervasive Computing and Situationally-Induced Im-pairments and Disabilities. *Proc. HCI Int*. 2003, 1298-1302.

[23] Stone, M. *A Field Guide to Digital Color*. 2003.

[24] Tufte, E. *Envisioning Information*. 10th ed., 1990.

[25] Viénot, F., Brettel, H., Ott, L., Ben M'Barek, A., Mollon, J. What Do Colour-Blind People See? *Nat.* 376, 1995, 127-128.

[26] Wakita, K., Shimamura, K. SmartColor: Disambiguation Framework for the Colorblind. *Proc. ASSETS 2005*, 158-165.

[27] Ware, C. *Information Visualization: Perception for Design*. 2000.

[28] Wyszecki, G., Stiles, W. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. 2nd ed., 2000.