

Journal of Electronic Imaging

SPIEDigitalLibrary.org/jei

Hue constrained matrix optimization for preferred color reproduction

Heng Zhang
Huaping Liu
Shuxue Quan



Hue constrained matrix optimization for preferred color reproduction

Heng Zhang
Huaping Liu

Oregon State University
School of Electrical and Computer Engineering
4115 Kelley Engineering Center
Corvallis, Oregon 97331-5501
E-mail: henghengzhang@gmail.com

Shuxue Quan

Broadcom Corporation
5300 California Avenue
Irvine, California 92617

Abstract. A method for preferred color reproduction with a psychophysical study based upon a real-life scene demonstrated application is presented. A color correction matrix optimization algorithm is first introduced which applies additional hue constraints on top of the CIE delta E magnitude error to emphasize the importance of hue in the subjective quality of color reproduction. An application is implemented by applying matrices optimized for the three dominant memory colors: skin, sky and green based on each pixel's categorical classification which is obtained through an explicitly defined boundary model. A psychophysical study was carried out to systematically investigate observers' hue preference in terms of the reproduction of individual memory colors as well as when multiple memory colors exist in the scene simultaneously. Conclusions suggest that observers' preference for individual memory colors is consistent with regard to scene composition. When a scene with multiple memory colors is evaluated, skin is given the highest priority in the overall preference scale, followed by sky and then green. In practice, the optimized preferred color reproduction may be achieved by first locating the matrices for individual memory colors through a rank order study, and then applying them based on each pixel's classification. © 2012 SPIE and IS&T. [DOI: [10.1117/1.JEI.21.3.033021](https://doi.org/10.1117/1.JEI.21.3.033021)]

1 Introduction

Among the many factors that can affect the subjective quality of an image, color reproduction is widely agreed to be a major one. Observers are able to judge the quality of the color reproduction of an image without knowledge of the original scene with the help of a capability developed from everyday life experiences and some memory colors such as human skin tone, blue sky and green vegetation are often used as important references during this evaluation process. Memory colors refer to certain colors associated with our daily life in such a way that observers can easily connect an object with a certain color in their memory. Memory colors have two important attributes: 1. these colors exist in fairly

consistent manners in nature; 2. they appear often enough such that every observer has his own memory for these colors. As just mentioned, skin tone, sky blue, and green are dominant memory colors and the quality of their reproduction can greatly affect the subjective quality of a reproduced color image.

As a common practice for digital image pipelines, color constancy is an essential and sometimes the only functional block in terms of color processing. Most of the color constancy algorithms consist of a two-stage process: during the first stage, the white point of the scene is estimated, and a diagonal matrix is applied to white balance the image; during the second stage, which is the concern of this paper, a color correction matrix is optimized for the specific image sensor, and the resulting matrix is applied to the white balanced image so that the data in the device color space can be transformed into a device independent standard color space. The optimization of this color correction matrix can be achieved by minimizing a cost function such as the color difference defined by CIE76, CIE94 or CIEDE2000 using the naïve least-squares regressions or with constraints such as the white point preserving least squares regressions and the weighted white point preserving least squares regressions.^{1,2} These methods focus on minimizing the magnitude error whose mathematical representation can vary from a simple equation like CIE76 ΔE_{ab}^* to a quite sophisticated form such as the CIEDE2000 ΔE_{00}^* . However, the rotational hue error, which can be of greater importance (demonstrated by the example on page 108),³ is left out of consideration. In a three-dimensional (3-D) color space defined by lightness, hue and chroma, it is intuitive thinking to include the rotational hue error into the matrix optimization process given the fact that color names such as red, yellow and green are defined along planes with constant hues and for the same amount of magnitude error, skin color can be reproduced in a reddish or a greenish tone where the latter case is visually objective. While this hue constrained matrix optimization method can be applied in a very general manner, we would like to explore its usage in preferred color reproduction in this work.

Paper 11299P received Oct. 25, 2011; revised manuscript received Jul. 13, 2012; accepted for publication Jul. 23, 2012; published online Aug. 16, 2012.

0091-3286/2012/\$25.00 © 2012 SPIE and IS&T

Memory colors and preferred color reproduction have been studied extensively in the literature and a plethora of work can be found in the patent industry dealing with specific implementations. Previous research through different methodologies,⁴⁻¹¹ has shown that memory colors of skin tone, sky blue, and foliage green do not have the same chromaticities as the original object colors especially when given image context such as sky, skin or plant present in the reproduced image. No strong correlation was found between the subjective quality of the color rendition and color accuracy and that significant hue shifts were found which in many cases are probably in the direction of the most impressive or pleasing chromatic attribute. Evidence has shown that the above colors are remembered with increased purity or saturation and the observation is consistent across different racial groups. These research results have been translated into specific implementations in practice with examples documented in Refs. 12–16. In Ref. 12 memory colors are processed using a coefficient group interpolated or extrapolated from multiple precalculated coefficient groups, or through a look up table. In this method, the degree of correction can be adjusted with ease and at least two of the hue, chroma and lightness signals are processed simultaneously. In Ref. 13 an apparatus is described for preferred color reproduction by first transforming the pixel values into lightness, chroma and hue, and then consistently and smoothly moving the hue values within a predefined region toward predetermined preferred colors or away from predetermined objectionable colors. This work also specifically defines preferred hue lines for certain patches on the Macbeth Color Checker. General methods are introduced without detailed mathematical descriptions.^{12,13} Other publications include methods such as the image-wide histogram manipulation based on the detected skin pixels,¹⁴ modeling skin color using Gaussian distributions in the $Yu'v'$ space and transform those pixels into the preferred color zone,¹⁵ or to find the memory color pixels in certain CbCr zones and apply some operator to these pixels to achieve preferred reproduction.¹⁶

This work contributes to the literature by explicitly building a link between color correction matrix optimization and preferred color reproduction. We also report results from a psychophysical study which leads to some general recommendation for the optimized preferred color reproduction of real-life scenes based upon the proposed hue constrained matrix optimization method. In Sec. 2, we give a statement of the problem by reviewing the color processing steps in digital image pipelines including the necessary color space transformations and demonstrate the importance of directional hue error in preferred color reproduction. In Sec. 3, the proposed method is discussed. In Sec. 4, we report some experimental results from a psychophysical study using rank order and paired comparison evaluations, followed by discussions and conclusions in Sec. 5.

2 Statement of the Problem

In this section, we review the color image formation process and the essential color space transformations that convert image data from the device dependent RGB color space to the approximately perceptually uniform CIELAB color space. Related reviews can be easily found in the public domain. We feel necessary to include it in this text for completeness and reference purposes. We will also briefly

review some of the existing techniques for color correction matrix optimization, and demonstrate the importance of the rotational hue error at the end of this section.

2.1 Color Image Formation Process

In color processing, a widely adopted model for describing the formation process of a color image is through the convolution of the spectra of the illuminant, the surface reflectance of the scene, and the sensitivity of the image acquisition system (in our case an image sensor) over the entire spectrum. For any specific pixel location (i, j) corresponding to the i^{th} row ($i = 1, 2, \dots, \text{NUM}_{\text{row}}$) and the j^{th} column ($j = 1, 2, \dots, \text{NUM}_{\text{col}}$) of an image, if λ represents the wavelength of the electromagnetic wave in nanometer, given an illuminant with a spectral power distribution $E(\lambda)$, the corresponding surface reflectance $R_{i,j}(\lambda)$, and the relative spectral sensitivity of the image sensor $S_{i,j}(\lambda)$, the resulting value of the pixel $P_{i,j}$ can be formulated as:

$$P_{i,j} = \int_{\lambda} E(\lambda) R_{i,j}(\lambda) S_{i,j}(\lambda) d\lambda, \quad (1)$$

where global illuminant is assumed. $S_{i,j}(\lambda) = S_k(\lambda)$ in which $k = 1, 2, \dots, \text{NUM}_{\text{cha}}$ (NUM_{cha} is the number of color channels of the sensor) depending on the arrangement of the pixel array. With the color filter array (CFA) arranged in Bayer pattern ($\text{NUM}_{\text{cha}} = 3$) as in most of the commercial image sensors, four values are obtained within each 2×2 quadrant among them one is red, one is blue, and two are green. This $\text{NUM}_{\text{row}} \times \text{NUM}_{\text{col}} \times 1$ array can be interpolated into to an $\text{NUM}_{\text{row}} \times \text{NUM}_{\text{col}} \times 3$ array such that every pixel has all three of red, green, and blue signals to form a full color image. In practice, the effective range of λ is limited by means such as infrared filters to around 400 to 700 nm in order to match the visible spectrum of the Human Visual System (HVS) whose characteristics are described by the International Commission on Illumination (Commission internationale de l'éclairage, CIE) color-matching functions.

In Fig. 1(a), we show the relative spectral sensitivity of the image sensor used for this study whose responses are denoted by $S_r(\lambda)$, $S_g(\lambda)$ and $S_b(\lambda)$ for red, green, and blue channels, respectively. In Fig. 1(b), the CIE color-matching functions are plotted and we denote them with $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$.

We will now rewrite Eq. (1) in matrix format for mathematical manipulability. Let N be the number of points from sampling the effective wavelength range (for example, $N = 31$ if we sample 400 nm to 700 nm for every 10 nm), and M be the number of reflective surfaces involved in the scene under consideration, then the following equation can be written:

$$\begin{bmatrix} \mathbf{R} \\ \mathbf{G} \\ \mathbf{B} \end{bmatrix}_D = \begin{bmatrix} \mathbf{S}_r \\ \mathbf{S}_g \\ \mathbf{S}_b \end{bmatrix} \cdot \text{diag}(\mathbf{E}) \cdot \mathbf{R}, \quad (2)$$

or in a more condensed form:

$$\mathbf{RGB}_D = \mathbf{S} \cdot \text{diag}(\mathbf{E}) \cdot \mathbf{R}, \quad (3)$$

where \mathbf{RGB}_D is the sensor's original response ($3 \times M$ matrix) with a subscript D for the notation of device, \mathbf{S} is the sensor's relative spectral sensitivity ($3 \times N$ matrix), $\text{diag}(\mathbf{E})$ is the

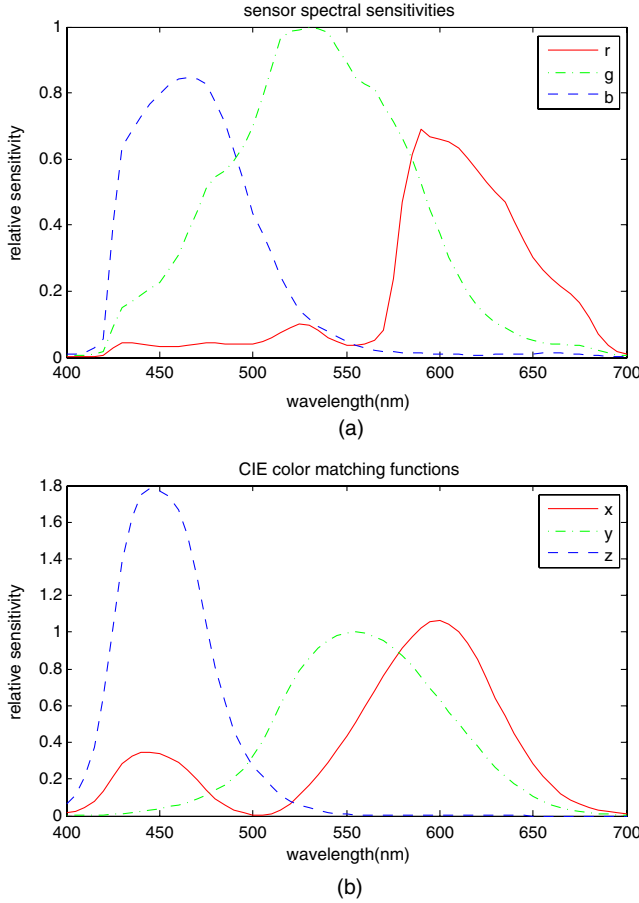


Fig. 1 (a) Spectral sensitivity response of the sensor used in this study; (b) CIE color-matching functions.

spectral power distribution of the illuminant ($N \times N$ matrix), and \mathbf{R} is the scene's spectral reflectance ($N \times M$ matrix).

Similarly, the tristimulus values for the same stimulus and reflectance surfaces can be written as:

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} \cdot \text{diag}(\mathbf{E}) \cdot \mathbf{R}, \quad (4)$$

or:

$$\mathbf{XYZ} = \mathbf{xyz} \cdot \text{diag}(\mathbf{E}) \cdot \mathbf{R}, \quad (5)$$

where \mathbf{XYZ} is the tristimulus values ($3 \times M$ matrix) and \mathbf{xyz} is the CIE color-matching functions ($3 \times N$ matrix).

2.2 Color Space Transformations

Equations (3) and (5) hint that if a perfect linear transformation exists through a 3×3 matrix between \mathbf{RGB}_D and \mathbf{XYZ} , an exact mapping could then be found between the sensor device and the HVS. Unfortunately, this mapping hardly exists and it can at most be optimized in the least-squared-errors sense.⁷ Researchers have also found that \mathbf{RGB}_D and \mathbf{XYZ} are not perceptually uniform color spaces. In practice, the device color space is converted into a standard color space such as sRGB with a transformation optimized for criteria based upon a perceptually more uniform color space such as CIELAB.

If \mathbf{RGB} is a $3 \times M$ matrix in the linear sRGB space, $\text{diag}(\mathbf{M}_{WB})$ is a 3×3 diagonal white balancing matrix, and \mathbf{M}_{CC} is a 3×3 color correction matrix which is the focus of this study, then the following equation can be written:

$$\mathbf{RGB} = \text{diag}(\mathbf{M}_{WB}) \cdot \mathbf{M}_{CC} \cdot \mathbf{RGB}_D. \quad (6)$$

While the actual processing within the digital pipeline may stop here, for matrix optimization purposes, we need to continue with the transformation from \mathbf{RGB} to \mathbf{XYZ} which is defined by:

$$\mathbf{XYZ} = \mathbf{M}_{709} \cdot \mathbf{RGB}, \quad (7)$$

and for D65:³

$$\mathbf{M}_{709} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7151 & 0.0721 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix}. \quad (8)$$

The above \mathbf{XYZ} values can now be transformed to the perceptually uniform CIELAB space as:

$$\begin{aligned} \mathbf{L}^* &= 116 \cdot f_{ab}(\mathbf{Y}/Y_n) - 16 \\ \mathbf{a}^* &= 500 \cdot [f_{ab}(\mathbf{X}/X_n) - f_{ab}(\mathbf{Y}/Y_n)] \\ \mathbf{b}^* &= 200 \cdot [f_{ab}(\mathbf{Y}/Y_n) - f_{ab}(\mathbf{Z}/Z_n)], \end{aligned} \quad (9)$$

where X_n, Y_n and Z_n are the tristimulus values of the reference white point, and

$$f_{ab}(t) = \begin{cases} t^{1/3}, & \text{if } t > (6/29)^3 \\ 1/3 \cdot (29/6)^2 t + 16/116, & \text{otherwise} \end{cases}. \quad (10)$$

The aim of the color correction matrix optimization process is to solve for \mathbf{M}_{CC} so that certain criteria can be satisfied. One example is to define a cost function which could be the average CIE76 ΔE_{ab}^* , CIE94 ΔE_{94}^* and CIEDE2000 ΔE_{00}^* of the test target patches used for the optimization and the condition for the optimization is to minimize this cost function.

In this work, we constrain the optimization process so that the hue of certain target colors can be precisely reproduced without overly worrying about the total magnitude error or perceptual uniformity of the modeling process. In order to improve the mathematical manipulability as well as for demonstration needs, we choose to use CIE76 ΔE_{ab}^* in the study and the color difference referred to hereafter will be ΔE_{ab}^* . We also used ΔE_{94}^* during the preliminary experiment stage for comparison purposes and the results were very similar. Given a reference set of $\mathbf{L}_{ref}^*, \mathbf{a}_{ref}^*$ and \mathbf{b}_{ref}^* , the color difference between the actual data $\mathbf{L}^*, \mathbf{a}^*$ and \mathbf{b}^* and the reference is defined as:

$$\begin{aligned} \Delta \mathbf{L}^* &= \mathbf{L}^* - \mathbf{L}_{ref}^* \\ \Delta \mathbf{a}^* &= \mathbf{a}^* - \mathbf{a}_{ref}^* \\ \Delta \mathbf{b}^* &= \mathbf{b}^* - \mathbf{b}_{ref}^* \\ \Delta E_{ab}^* &= \sqrt{\Delta \mathbf{L}^{*2} + \Delta \mathbf{a}^{*2} + \Delta \mathbf{b}^{*2}}. \end{aligned} \quad (11)$$

2.3 Previous Work

Existing techniques for the optimization of color correction matrix include the naïve least-squares (LS) regressions which minimizes ΔE_{ab}^* , white-point preserving least-squares (WPPLS) regressions¹ which is a constrained form of LS with additional requirement that the sum of each row of the matrix is one, weighted white-point preserving least squares (WTWPPLS) regressions which increases the weight of certain patches of the test target for more visually pleasing colors.²

Generally applicable to the above mentioned or other algorithms as well as the topic discussed in this work, the transformation between the device data \mathbf{RGB}_D and the tristimulus data \mathbf{XYZ} including linear constraints can be modeled linearly, and a solution can be analytically derived. Taking the simplest naïve least-squares method as an example, given the reference tristimulus $\mathbf{XYZ}_{\text{ref}}$, the problem can be described as finding a solution to minimize the following cost function:

$$f_{\text{LS}} = \|\mathbf{XYZ}_{\text{ref}} - \mathbf{XYZ}\|^2, \quad (12)$$

where $\|\cdot\|$ is the $L2$ norm of (\cdot) and

$$\mathbf{XYZ} = \mathbf{M}_{709} \cdot \mathbf{M}_{\text{WB}} \cdot \mathbf{M}_{\text{CC}} \cdot \mathbf{RGB}_D = \mathbf{M} \cdot \mathbf{RGB}_D, \quad (13)$$

in which $\mathbf{M} \doteq \mathbf{M}_{709} \cdot \mathbf{M}_{\text{WB}} \cdot \mathbf{M}_{\text{CC}}$ is the target matrix to be solved for without considering chromatic adaptation.

Taking the first derivative of Eq. (12), set it to zero and use the substitution in Eq. (13), we can arrive at the following equations:

$$\begin{aligned} \mathbf{M}^T \cdot (\mathbf{XYZ}_{\text{ref}} - \mathbf{M} \cdot \mathbf{RGB}_D) &= 0 \\ \mathbf{XYZ}_{\text{ref}} &= \mathbf{M} \cdot \mathbf{RGB}_D \\ \mathbf{XYZ}_{\text{ref}} \cdot \mathbf{XYZ}_{\text{ref}}^T &= \mathbf{M} \cdot \mathbf{RGB}_D \cdot \mathbf{XYZ}_{\text{ref}}^T, \end{aligned} \quad (14)$$

where the superscript T denotes the transpose of a matrix, and the final solution is:

$$\mathbf{M} = (\mathbf{XYZ}_{\text{ref}} \cdot \mathbf{XYZ}_{\text{ref}}^T) \cdot (\mathbf{RGB}_D \cdot \mathbf{XYZ}_{\text{ref}}^T)^{-1}. \quad (15)$$

The solution found by Eq. (15) will probably fail to minimize the color difference in a perceptually uniform color space since it is at best optimized on a least-squares sense. Nonetheless, it is a good initial value for the numerical approach that follows in the optimization process. More advanced analysis in the linear domain can be found.^{17,18}

2.4 Taking Directional Hue Error into Account

The above discussed constrained or unconstrained methods utilize magnitude errors as their metrics while discarding any information about the directional hue error, which could be of much greater importance than the magnitude error especially when preferred color reproduction is concerned. In Fig. 2, we plot two color samples: skin and sky on the a^*b^* coordinates with $L^* = 65$. The target a^* and b^* values are calculated using the light skin and blue sky patches from the Macbeth color checker under D65 and the circles have a radius of eight. The just noticeable difference in the CIELAB space is around 2.3 (Chapter one, page 31),¹⁹ from a practical point of view as well as for better distinguishability, we choose a radius of eight in the figure.

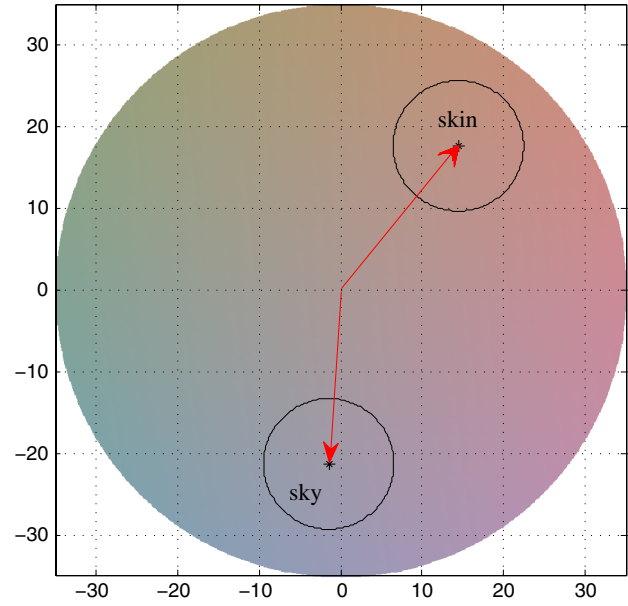


Fig. 2 Target skin and sky colors on a^*b^* coordinates at $L^* = 65$. The circles have a radius of 8.

This figure shows that depending on the directional hue error, with the same magnitude of color difference from the target, skin can vary from reddish to yellowish to greenish, and sky can vary from cyan to blue to purple. Given the range of possible variations, observers may have quite different opinions about the quality of the color reproduction of the image.

3 Proposed Method

To address the above directional hue error and preferred color reproduction issue, we propose a hue constrained method for the optimization of color correction matrix based upon the white-point preserving method which can be described by a constrained optimization problem as finding a set of parameters represented by vector \mathbf{x} such that the cost function $f(\mathbf{x})$ is minimized:

$$\arg \min_{\mathbf{x}} f(\mathbf{x}) = \varepsilon(\Delta L^*, \Delta a^*, \Delta b^*) \quad (16)$$

subject to:

$$\text{cneq}(\mathbf{x}) < 0 \quad \text{ceq}(\mathbf{x}) = 0, \quad (17)$$

where $\mathbf{x} = [x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32}]$ is a vector containing elements of the 3×3 matrix \mathbf{M} defined in Eq. (18):

$$\mathbf{M} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}, \quad (18)$$

and \mathbf{M} is constrained by the white-point preserving condition:

$$\begin{aligned} x_{13} &= 1 - x_{11} - x_{12} \\ x_{23} &= 1 - x_{21} - x_{22} \\ x_{33} &= 1 - x_{31} - x_{32}. \end{aligned} \quad (19)$$

In Eq. (16), $\Delta \mathbf{L}^*$, $\Delta \mathbf{a}^*$, $\Delta \mathbf{b}^*$ are the difference vectors between the actual data set and the reference set defined by Eq. (11), $\varepsilon(\cdot)$ is the definition of the optimization criteria which can take different formats. In this study, we use ΔE_{ab}^* for a 24-patch Mather Color Checker and:

$$\varepsilon(\Delta \mathbf{L}^*, \Delta \mathbf{a}^*, \Delta \mathbf{b}^*) = \frac{\sum_{i=1}^{24} \sqrt{\Delta L_i^{*2} + \Delta a_i^{*2} + \Delta b_i^{*2}}}{24}, \quad (20)$$

where the subscript i represents the i^{th} ($i = 1, 2, \dots, 24$) patch of the standard test target.

Before introducing the general definition of Eq. (17), the following terms are needed: rotational hue error, chroma error and lightness error. The rotational hue error Δh is the difference in degrees between the actual hue h and the reference hue h_{ref} :

$$\Delta h = h - h_{\text{ref}}, \quad (21)$$

in which hue angle is limited to be between 0 deg and 360 deg. Let us denote $h_{\text{int}} = \tan^{-1}(b^*/a^*)$ whose unit is radians then

$$h = \begin{cases} h_{\text{int}}/\pi \cdot 180, & \text{if } h_{\text{int}} \geq 0 \\ (2\pi + h_{\text{int}})/\pi \cdot 180, & \text{if } h_{\text{int}} < 0 \end{cases} \quad (22)$$

Chroma error Δc is defined as the difference between the actual and the reference chroma values:

$$\Delta c = c - c_{\text{ref}} = \sqrt{a^2 + b^2} - \sqrt{a_{\text{ref}}^2 + b_{\text{ref}}^2}, \quad (23)$$

and lightness error is:

$$\Delta L = L - L_{\text{ref}}. \quad (24)$$

The directions of the rotational hue error and chroma error are demonstrated in Fig. 3 using the light skin patch as an example where the lightness error is not plotted since its meaning is more straightforward. In this figure, the straight blue line defines a constant hue along which the rotational hue error is 0°, and the blue arc defines a constant chroma line along with the chroma error is zero. The signs of the rotational hue error and chroma error are marked correspondingly on both sides of each of the constant lines.

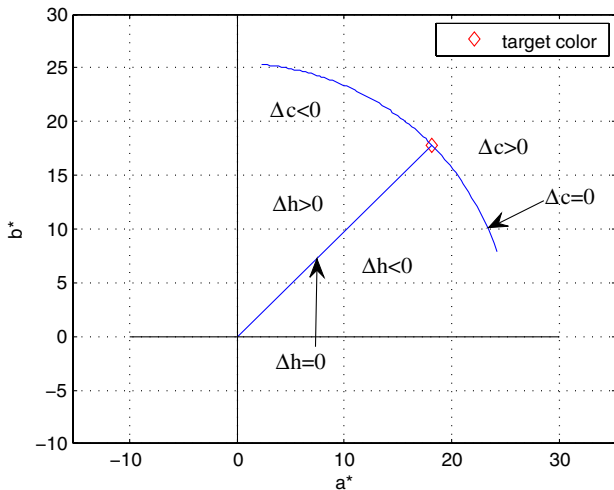


Fig. 3 Directional hue error and chroma error.

Equation (17) represents a set of inequality conditions $\text{cneq}(\mathbf{x}) < \mathbf{0}$ and a set of equality conditions $\text{ceq}(\mathbf{x}) = \mathbf{0}$. For example, if we would like to constrain the hue error of the light skin patch ($i = 2$) to be equal to -4° , the chroma error of the blue sky patch ($i = 3$) to be positive, and the lightness error of the foliage patch ($i = 4$) to be equal to zero, the following can be written:

$$\begin{aligned} \text{cneq}(\mathbf{x}) &\doteq [-\Delta c_3] < [0] \\ \text{ceq}(\mathbf{x}) &\doteq \begin{bmatrix} \Delta h_2 + 4 \\ \Delta L_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned} \quad (25)$$

Equations (16) and (17) define a constrained nonlinear optimization problem and one of the solutions can be found using the sequential quadratic programming (SQP) algorithm. Descriptions of the SQP algorithm can be found in Refs. 20–22 and its general method can be summarized as follows (according to the Matlab optimization toolbox help document by Mathworks):

Given the objective function f , the nonlinear inequality constraint vector cneq and nonlinear equality constraint vector ceq , the Lagrangian is defined as:

$$L = f + \sum_j \xi_j \text{cneq}_j + \sum_k \xi_k \text{ceq}_k, \quad (26)$$

where ξ is the Lagrange multiplier and its subscripts j and k are determined by the number of constraints. Equation (26) can be represented in a simplified format as:

$$L = f + \sum_j \xi_j \cdot g_j(x), \quad (27)$$

whose Hessian is the second-order derivative of the Lagrangian:

$$H = \nabla^2 L = \nabla^2 f + \sum_j \xi_j \nabla^2 g_j. \quad (28)$$

At each major iteration k , a quasi-Newton approximation of the Hessian is calculated using the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method as:

$$H_{k+1} = H_k + \frac{q_k q_k^T}{q_k^T s_k} - \frac{H_k^T s_k^T s_k H_k}{s_k^T H_k s_k}, \quad (29)$$

where

$$s_k = x_{k+1} - x_k \quad q_k = [\nabla f(x_{k+1})]^+. \quad (30)$$

A quadratic programming (QP) subproblem can be obtained by linearizing the nonlinear constraints as:

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \frac{1}{2} d^T H_k d + \nabla f(x_k)^T d \\ \text{subject to} \quad & \nabla g_j(x_k)^T + g_j(x_k) = 0, \quad j = 1, \dots, m_e \\ & \nabla g_j(x_k)^T + g_j(x_k) \leq 0, \quad j = m_e, \dots, m, \end{aligned} \quad (31)$$

which can be solved by a quadratic algorithm such as the interior point method²³ or the trust region reflective method²⁴ and its solution can be used for a new iteration until the constraints are met.

4 Experiments

In this section, we report the results from a psychophysical study including a rank order experiment and a forced choice paired comparison experiment in order to systematically investigate observer preference for rotational hue error in terms of individual memory colors as well as how the different colors interact within the same scene to arrive at a recommendation for the optimized overall reproduction.

4.1 Test Scenes

Hue constraints applied during the matrix optimization process will affect the overall color accuracy as reported in Ref. 25 and can also lead to some trade-offs between saturation and noise. In order to eliminate any possible influence

from saturation and noise trade-offs,²⁶ only carefully selected scenes with sufficient lighting are used in this study. To better isolate the different aspects of the problem and stay focused, we limit the study to the preference for memory colors and try to make other aspects such as artifacts from chromatic aberration not visually detectable. We also assume that among the dominant memory colors, sky and vegetation can mostly be found outdoors, whereas skin can be seen both outdoors and indoors.

Figure 4 depicts eight of the eleven scenes used during the study. Images not listed in the figure are indoor images taken in an office with neutrally colored background, one with a male Caucasian as the main object and some colorful objects in the background (named “Caucasian” for later references), one with the same background except with a male Asian as

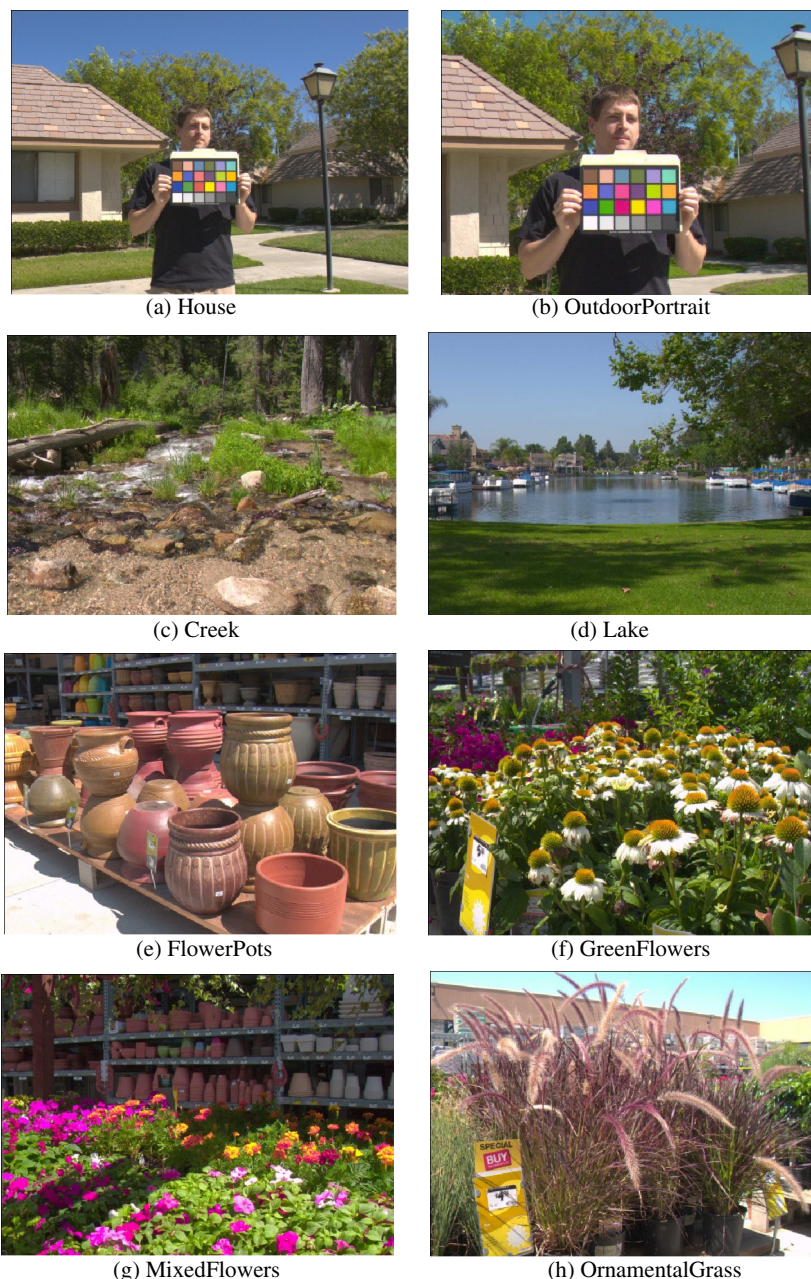


Fig. 4 Partial list of the test scenes.

the main object (named “Asian”), and the last one has the colorful objects in the previous two scenes except they are now the main objects (named “Toys”).

In the above scenes, the “House” image has all three of skin, sky, and green colors as well as some skin-colored artificial object (the roof). The “OutdoorPortrait” image is a cropped version of the “House” image in which we intend to increase the proportion of the skin and see how this change affects observers’ preferences. The “Creek” image has relatively rich textures and the only memory color in the scene is the different shades of green. The “Lake” image is mainly composed of sky and grass, with some skin-colored roof in the background. The “FlowerPots” scene is consisted of purely artificial objects in which some of the pots may fall into the category of skin color. The “GreenFlowers” scene has different shades of green and some colorful flowers with rich textures. The “MixedFlowers” scene is rich in texture with bright-colored flowers. The “OrnamentalGrass” scene has shades of green in both of the foreground and the background and the color of the grass is not easy for observers to recall from their memory.

All of the above scenes have a resolution of 1852×1324 except the “OutdoorPortrait” scene which is cropped to 1200×900 .

4.2 Test Setup and Procedure

The evaluations were performed in a dark room with a EIZOCOLOR CG241W high-end monitor. All observers have normal color vision and they were asked to adjust the seating until their eyes were aligned with the center of the monitor before the evaluation starts. The viewing distance is about 85 cm.

A full length evaluation includes a rank order experiment followed by a paired comparison experiment. Before the full

experiment was executed, a pilot rank order study was first carried out to assist the selection of constraints.

4.3 Rank Order Study

The rank order study was performed in a few steps. The first step was a pilot experiment using the “House” image. With the Munsell color checker in the scene, we used the light skin patch (patch 2), the blue sky patch (patch 3) and the foliage patch (patch 4) as target colors for skin, sky, and green, respectively. Through varying the hue constraint for each patch from -12° to 12° with a 2° step, 13 matrices were calibrated for each color and they were applied globally to the image, creating 13 images numbered from 1 to 13 correspondingly. With these three sets of images, the skin varies from a sunburned red to a very sickly yellow, the sky varies from a very greenish cyan to a purple, and the grass varies from a dry-looking yellow to a very plastic-like bluish green. By removing some of the variations that are overly visually objective, only nine images for each color were kept for the evaluation: images one through nine for skin and green and images two through 10 for sky.

In the pilot study, four subjects (two Chinese, one Korean, and one US born Caucasian; among them three are naïve observers and one is the main author) were asked to do a rank-order evaluation for the three sets of images. For each set to be displayed, the images were adjusted such that the area of interest for all images can be displayed on the same screen as demonstrated in Fig. 5 for the skin set. The observers were given the following instructions:

“In this experiment, you need to evaluate the quality of the designated areas (skin/sky/green) in terms of the quality of their color reproduction. Based on your everyday experience, please give the order of the images from the one you PREFER the most to the least.”

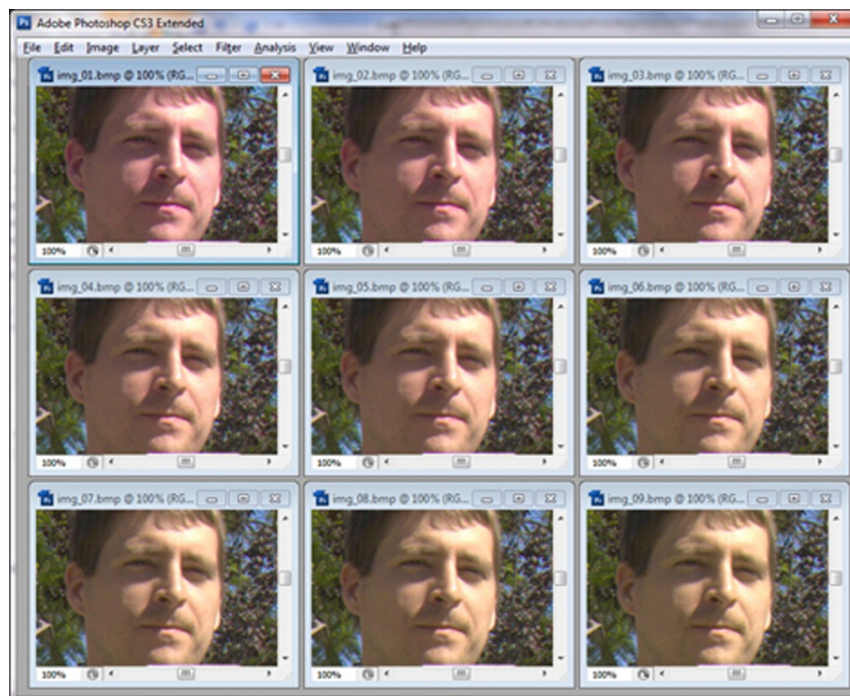


Fig. 5 Rank-order experiment screenshot.

Assume NUM_{obs} observers participated in the experiment using NUM_{sti} stimuli then a $NUM_{obs} \times NUM_{sti}$ matrix \mathbf{r} can be constructed to record the ranks of all stimuli given by the observers. The element located at the i^{th} ($i = 1, 2, \dots, NUM_{obs}$) row and the j^{th} ($j = 1, 2, \dots, NUM_{sti}$) column of \mathbf{r} corresponds to the rank given by the i^{th} observer to the j^{th} stimulus. For example, in a 4×9 matrix, a five at the 2nd row and the 3rd column means that observer two ranked the 3rd sample the 5th place during the test. We can then compute a raw score matrix \mathbf{s}_{raw} like this:

$$\mathbf{s}_{raw}(i, j) = NUM_{sti} - \mathbf{r}(i, j), \quad (32)$$

and the rank score vector $\mathbf{s} = [s_1, s_2, \dots, s_{NUM_{sti}}]$ can be calculated as:

$$S_j = \frac{\sum_{i=1}^{NUM_{obs}} [NUM_{sti} - \mathbf{r}(i, j)]}{NUM_{obs} \cdot (NUM_{sti} - 1)}, \quad j = 1, 2, \dots, NUM_{sti}. \quad (33)$$

With Eq. (33), the rank score of each stimulus is normalized to between zero and one and a higher score indicates that the stimulus is ranked as having better quality for the given criteria. In the extreme cases, a score of 1 indicates that the stimulus was unanimously ranked as the best and a score of zero means that the stimulus was unanimously ranked as the worst. The rank scores from this pilot study were calculated and analyzed but the details of it are left out of this discussion since the size of the observer pool is too small for any conclusions. The results from this pilot study, however, clearly indicate that the observers can distinguish the differences among the stimuli and do have their own preferences.

After the above pilot experiment, a full length evaluation was carried out during which each observer was asked to do the rank order experiment first, followed by a forced choice paired comparison experiment. The following text introduces the details of the rank order study and the paired comparison study will be discussed in a later section. A total of 14 subjects (one expert who is the main author and 13 naïve observers) participated in this full test including the ones in the pilot study, among them ten are Chinese (four females one in her 30's, three in their 40's; six males, four in their 30's and two in their 40's). One is a Korean male in his 30's and three are US-born Caucasian males in their 30's. One of the Chinese male in his 40's was too extreme in his evaluation and was noticeably inconsistent during the evaluation therefore his results were excluded from the later data analysis. It is not the intention to study how age, gender or ethnic origin affect the preference for memory color reproduction in this study so all of the data will be analyzed without distinguishing each observer's biological background.

The rank scores of the "House" scene are plotted in Fig. 6 below. As most of the observers described that they prefer the reproduction to be "natural" or "real" from this figure, we can tell clear trends of their preferences regardless of the descriptions. For skin tone, image seven corresponds to 0° of hue error and a negative error (images one to six) means a more reddish reproduction. This figure shows that a mild reddish reproduction for skin is preferred and too much yellow deteriorates the subjective quality of the reproduction. For blue sky, it is obvious that a mild

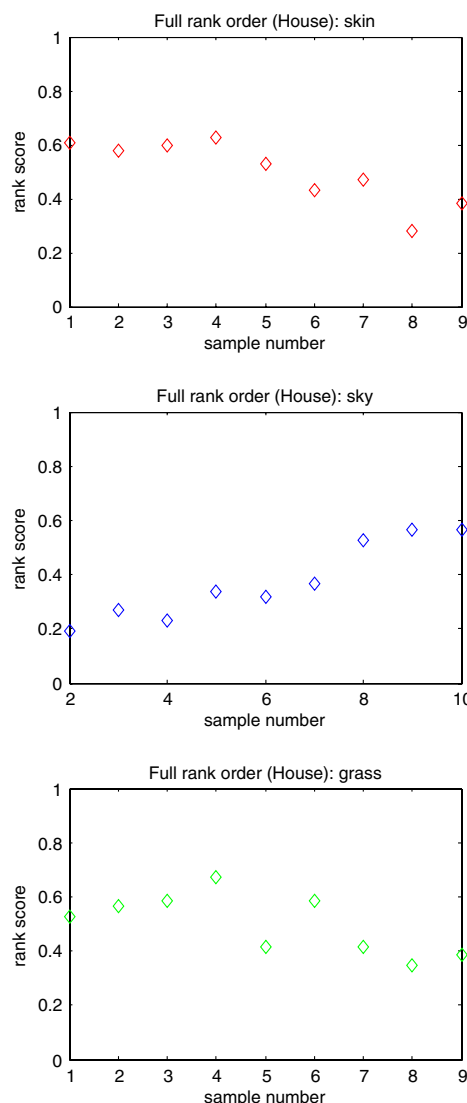


Fig. 6 Full rank order test of the "House" scene.

positive hue error makes the most preferred reproduction and too much green lowers the reproduction quality by making the observers think that it is "fake" and "nonrealistic." Grass color is more preferred to have some mild negative hue error which makes the grass look more vivid and sufficiently watered and too much blue can result in a plastic look in the rendered image which is not preferred. We expected the scores to be monotonically decreasing on both sides of the peak but the actual plots are slightly different. Possible reasons may include how the images were arranged on the screen, the relatively small difference between consecutive variations and the relatively small size of the subject pool.

For the full length experiment, two additional indoor images ("Caucasian" and "Asian") were included for the skin tone preference evaluation and the rank results are shown in Fig. 7 in which the scores from the "House" scene are also plotted. This figure indicates that observers have very similar preference for Caucasian and Asian skin tones in terms of rotational hue error. The curve for the Asian skin is smoother than that for the Caucasian skin

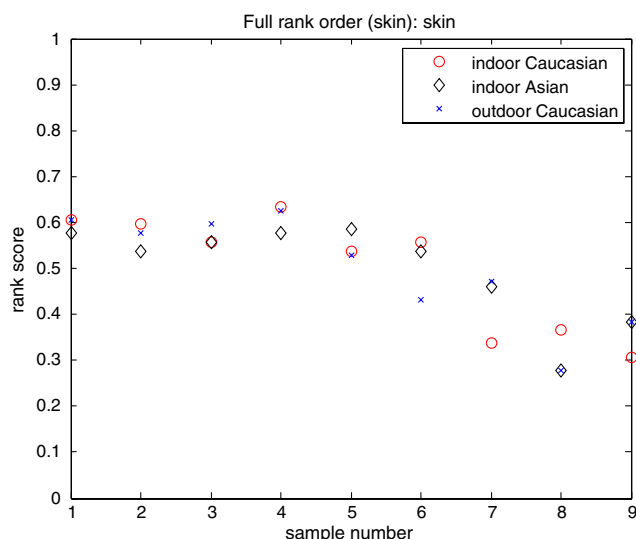


Fig. 7 Rank scores of skin reproduction from multiple scenes.

probably because the Asian skin has a lower overall reflectance level and the difference caused by the same amount of rotational hue error is perceived at a lesser degree than that of the Caucasian skin.

We also would like to study whether the observers' preference can be affected by the image context. To do so, with similar set up described earlier for the "House" scene, a different resizing scale was used so that more contexts around the color of interest can be seen by the observers as shown in Fig. 8. This experiment was included in the full length experiment for 10 observers (one expert who is the main author and nine naïve observers) and nine data sets are considered as valid (invalid dataset was described earlier). The results from this experiment are plotted in Fig. 9 which also

shows the scores from the test using the same images but displayed with fewer contexts.

Figure 9 shows that the basic trends of observer preference do not vary with image context despite the relatively small observer pool. The correlation coefficients between the scores from scenes with fewer and more contexts are 0.8791, 0.8997, and 0.8769 for skin, sky, and grass, respectively. With a sufficiently large observer pool, we would expect to see less fluctuation in the scores and better correlated curves. In this experiment, we found that although clearly instructed to focus on the color of interest, observers tend to use the surroundings and other factors such as contrast to assist their decision making process especially when subtle differences need to be distinguished which could also be a cause for the fluctuations.

4.4 Applying Multiple Matrices to a Single Image

In the above rank order study, each image was color corrected globally with a matrix optimized with hue constraints set for a single color. A global color correction offers its advantage in the simplicity of operation while the overall chromatic accuracy is compromised. When multiple colors need to be manipulated within the same scene, one way is to try to add more constraints during the optimization process and solve for a global matrix. However, this possibility did not prove itself to be the best option during the study since too many constraints could cause convergence problem or compromise the overall color accuracy to an unacceptable degree.

We instead take a different approach for the simultaneous manipulation of multiple colors. In a previous article,²⁷ the authors discussed a pixel clustering technique using raw image data, which could be used for the detection of memory color pixels and further developed to be applicable for the

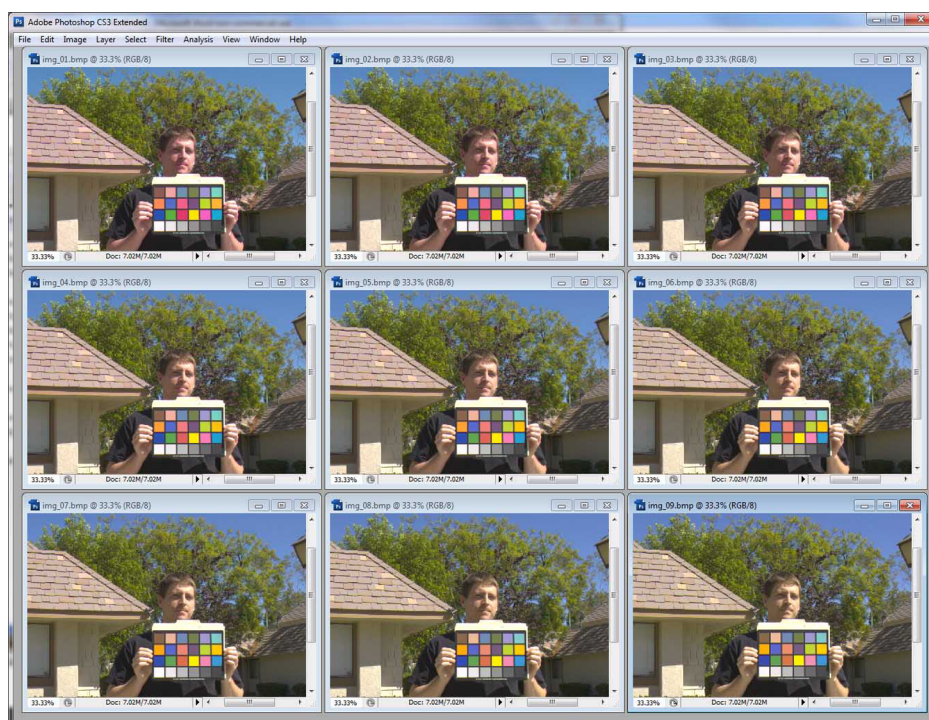


Fig. 8 Screen shot for rank order test with more contexts.

sRGB space. In this work, a simpler approach,¹⁶ is taken which does not rely on device dependent parameter calibration processes and detects memory color pixels in the YCbCr space that can be conveniently calculated from the sRGB data. Assuming a pixel $\mathbf{rgb}_{WB} = [r, g, b]^T$ is a 3×1

vector representing the three color channels of the pixel after white balancing, \mathbf{M}_{minDE} , \mathbf{M}_{skin} , \mathbf{M}_{sky} and \mathbf{M}_{green} are 3×3 matrices optimized for the best color accuracy, skin,

sky and green, respectively. We then process this pixel with the following procedure:

$$\mathbf{rgb}_{CC} = [r_{CC}, g_{CC}, b_{CC}] = \mathbf{rgb}_{WB} \bullet \mathbf{M}_{minDE}, \quad (34)$$

$$C_r = r_{CC} - g_{CC} \quad C_b = b_{CC} - g_{CC}, \quad (35)$$

$$\text{colortype} = \begin{cases} \text{green,} & \text{if } C_b < C_r \text{ \& } C_b > 10 \cdot C_r \\ \text{sky,} & \text{if } C_b < -0.4 \cdot C_r \text{ \& } C_b > -0.6 \cdot C_r \\ \text{skin,} & \text{if } C_b < -0.1 \cdot C_r \text{ \& } C_b > -1.25 \cdot C_r \text{ \& } r_{CC} < 1.75 \cdot g_{CC} \\ \text{other,} & \text{otherwise} \end{cases} \quad (36)$$

$$\mathbf{rgb} = \begin{cases} \mathbf{rgb}_{WB} \cdot \mathbf{M}_{green}, & \text{if color type is green} \\ \mathbf{rgb}_{WB} \cdot \mathbf{M}_{sky}, & \text{if color type is sky} \\ \mathbf{rgb}_{WB} \cdot \mathbf{M}_{skin}, & \text{if color type is skin} \\ \mathbf{rgb}_{WB} \cdot \mathbf{M}_{minDE}, & \text{if color type is other} \end{cases} \quad (37)$$

where \mathbf{rgb} is the final reproduction of the pixel.

Figure 10 depicts the detected memory color pixels (highlighted pixels) of the “House” scene obtained through the above procedure with an additional 5×5 Gaussian filter applied for noise suppression. This procedure offers a reasonably good detection rate with some artificial objects falling into the memory color categories as shown in the example of Fig. 10 in which the tiles on the roof are detected as skin color. More advanced techniques such as face detection can certainly be exercised to improve the accuracy of the pixel detection process but it is a topic beyond the scope of this study and the possible side effects of false positive detection are not studied in this work.

4.5 Paired Comparison Experiment

A forced choice paired comparison experiment was carried out to study how the observer preference is affected when multiple memory colors present in the image simultaneously. Based on the pilot rank order study, two matrices were chosen for each memory color and a total of eight combinations can be obtained for each scene using the procedure described in Sec. 4.3 with exception of the indoor scenes for which only skin color is taken into account and 4 combinations were created for each scene. For each of the scenes, an additional image corrected using the matrix without any hue constraint is also included. We would like to study observers’ sensitivity to subtle differences when the considered memory color is put in context of the complete scene and the two indoor matrices were chosen to have constraints that are close to each other. The constraints used for these matrices are summarized in Table 1.

In order to reduce the length and therefore observers’ stress level from the full evaluation which included the rank order experiment discussed earlier and the paired comparison experiment, the sky variation for all other outdoor scenes were dropped except for the “House,” “OutdoorPortrait,” and the “Lake” scenes. The “Lake” scene only has some skin colored pixels on the roofs of the houses in the background

which do not occupy a big enough portion to attract the observers’ attention in the evaluation, we could have chosen to drop its skin variation but decided to keep it in order to check the observers’ consistency when those skin variations just serve as redundant stimuli. At the end, 3 scenes were processed with nine different manipulations, and eight scenes were processed with five different manipulations, resulting in a total of 188 pairs of images for comparison.

The average time for an observer to complete the random and paired comparison tests is about 60 min with the shortest being 40 min and the longest being 100 min. During the experiment, the observers were encouraged to describe what they pay attention to for each type of scene, and how their priorities were set when making a decision. They were also encouraged to describe freely about their feelings of the images. This evaluation puts the observers under a certain stress level but most of them managed to handle it and were able to reduce the stress and improve their focus by describing the decision-making process to the main author who sat in the same room the full time during the evaluation.

We implemented a graphical user interface (GUI) to display the image pairs and record the observers’ preferences. At the beginning of the paired comparison experiment, the instruction is given as follows:

“In this experiment, images will be presented to you in pairs. Based on your everyday experience, please select the one that you think has a better color reproduction quality. Please ignore any other aspects rather than color. If it’s difficult to make a decision, you are encouraged to select intuitively and you MUST make a choice.”

The GUI then starts to display the images in pairs, and records the observer’s choices in the background. All pairs of a scene are displayed before the GUI starts on the next scene. Within each scene, the order of the pairs displayed and the locations of the images on the GUI (left or right) are randomized. Between consecutive pairs, images of white noise are displayed for 1.5 s to reduce the effect of iconic memory. The background of the GUI is 20% gray converted to sRGB and Fig. 11 is a screenshot of the GUI:

For each scene, a $N \times N$ matrix is constructed to record each observer’s preference where N is the number of different manipulations for the scene. If we denote this matrix as C_k for observer k then a one at $C_k(i, j)$ represents that the observer selected image j as his preferred reproduction for

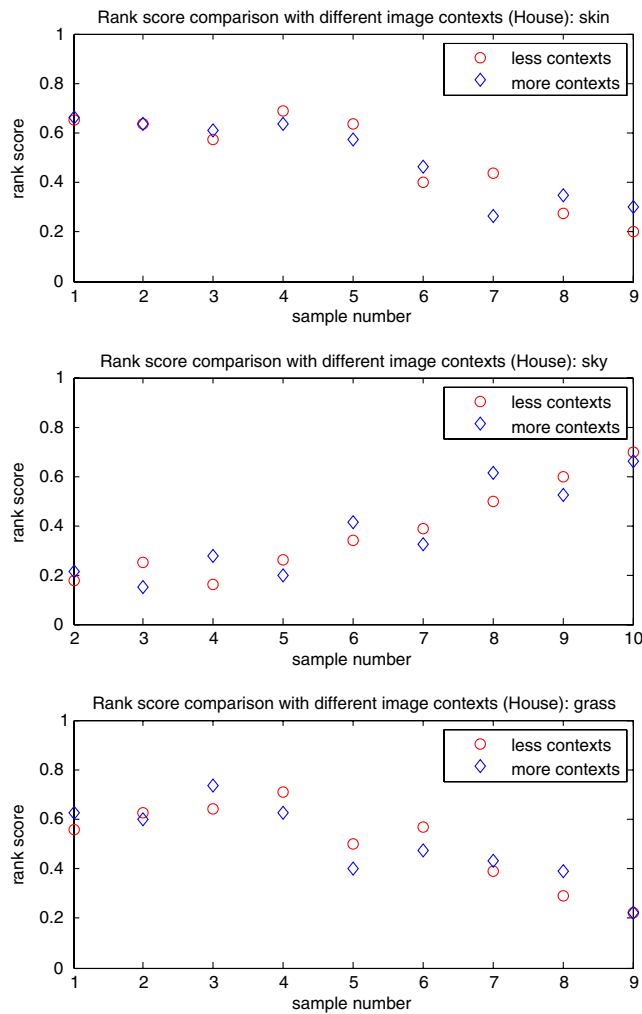


Fig. 9 Comparison of rank scores with different contexts.

the comparison between stimuli i and j . A 0.5 is assigned along the diagonal of the matrix since an image is never compared against itself and if done so we expect the preference to be purely random given sufficient trials. For each scene, a probability matrix \mathbf{P} can be calculated by dividing the sum of the matrices from all observers by the number of observers K :

$$\mathbf{P}_{i,j} = \frac{\sum_{k=1}^K \mathbf{C}_k(i,j)}{K}, \quad i, j = 1, \dots, N. \quad (38)$$

The quantified interval scales from this experiment can then be calculated using the inverse cumulative distribution function. To handle singular points caused by unanimously agreed selections, an approach suggested in Ref. 14 is taken which breaks the whole matrix into overlapping matrices that do not contain singular points and align the overlapped data points to form the final scales. The 95% confidence error bar plots of the scales are shown in Fig. 12 where for better reference, the data points are coded with letters: skin is represented by “S” or “s”, blue sky is represented by “B” or “b” and green by “G” or “g”. An upper case letter indicates that the correction for that color corresponds to the more preferred matrix whereas a lower case letter indicates that the less preferred matrix is used for the correction of a certain

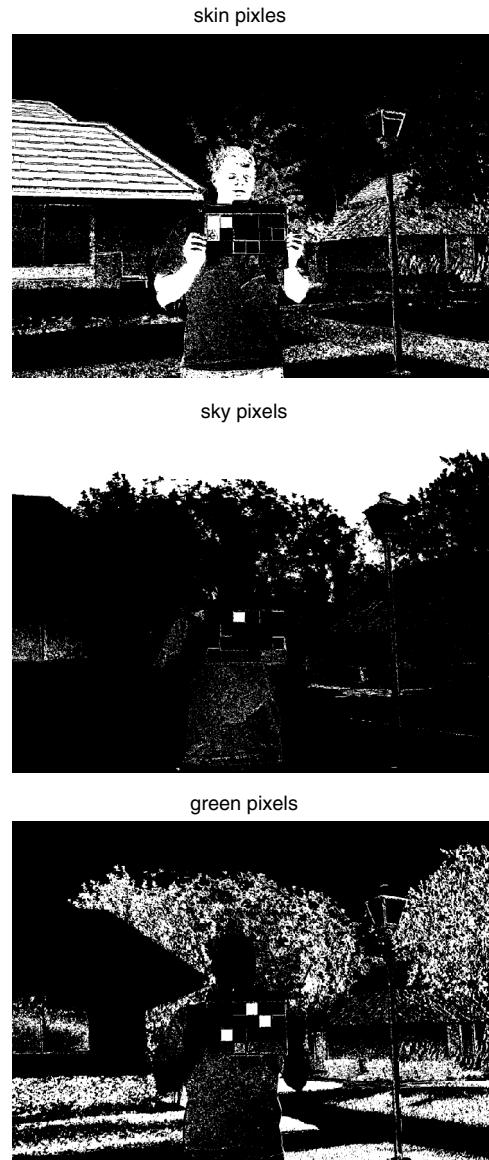


Fig. 10 Detected memory color pixels.

Table 1 Matrix constraints used for the paired-comparison stimuli.

Outdoor	Better	Worse
Skin	$\Delta h = 0^\circ$	$\Delta h = -6^\circ$
Sky	$\Delta h = -4^\circ$	$\Delta h = 2^\circ$
Green	$\Delta h = -6^\circ$	$\Delta h = 0^\circ$
Indoor	Better	Worse
skin	$\Delta h = -4^\circ$	$\Delta h = -8^\circ$

color. For example, “SbG” represents that the image is corrected with more preferred matrices for skin and green and less preferred matrix for sky, and “none” is used to indicate images corrected with matrix with no hue constraints. This notation system is used for subplots (a) through (h).

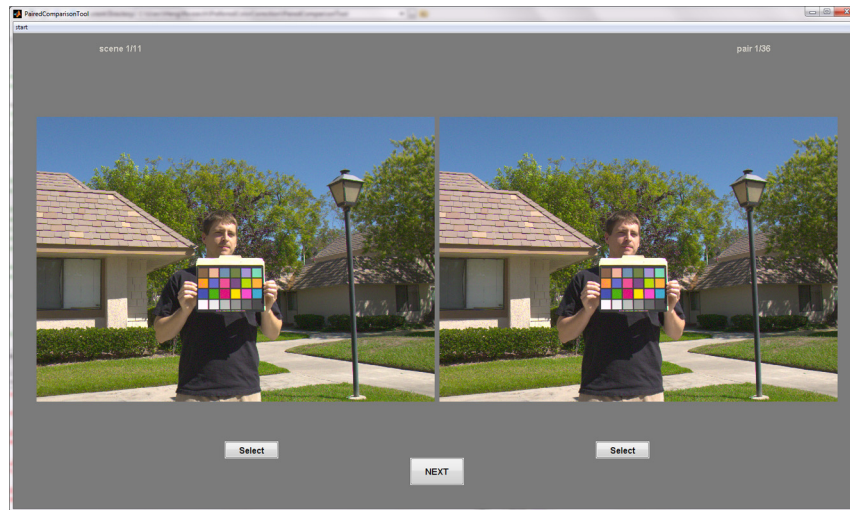


Fig. 11 Screen shot of the paired comparison GUI.

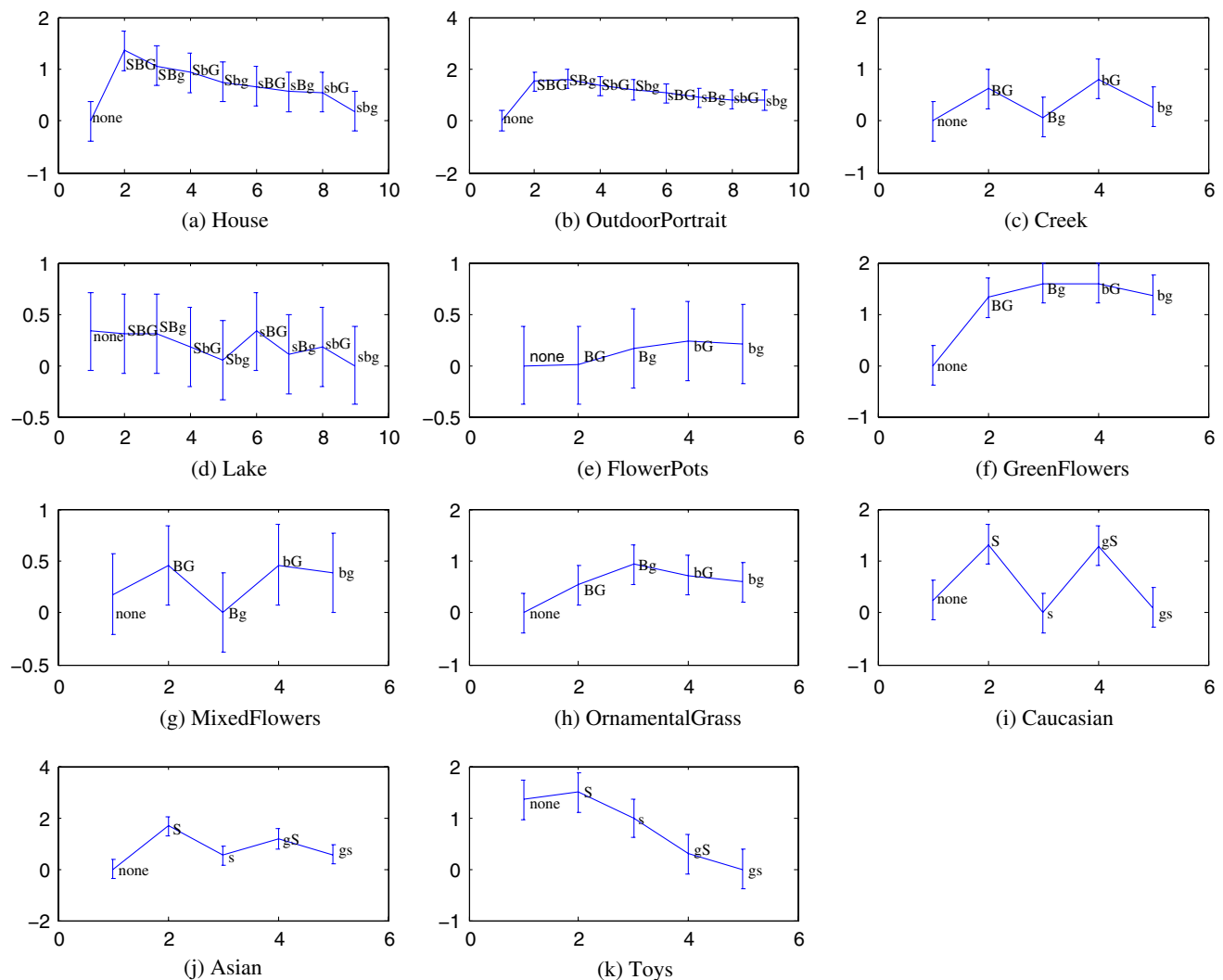


Fig. 12 The calculated z-scores from the forced choice paired comparison experiment.

From Fig. 12(a) for the “House” image, we can observe that the group with more preferred skin matrix is more preferred compared to the group corrected with the less preferred skin matrix, indicating that observers give their highest priority to the reproduction of skin when it presents in the foreground while the color reproduction of the other parts is not unpleasant enough to overrule their opinion. We can also tell that sky was given higher priority compared to green since there are small patches of grass scattered in the background while the sky occupies a significant portion in a continuous manner. There are also different shades and many details including shadows in both of the grass and foliage, which make the green areas less important than the sky area. The unconstrained matrix happened to reproduce the image in which the skin tone has too much red in it and was not surprisingly ranked as the least preferred. In Fig. 12(b) for the “OutdoorPortrait” image, the same trends hold except that the effect of green variation is smaller since the area of grass is now even smaller, while some observers were still able to distinguish the subtle differences in the background, some of them just randomly made their selections for those comparison pairs. For the “Lake” scene, the unconstrained matrix happened to reproduce the sky and grass in a manner that is very similar to stimulus one, as we can see in Fig. 12(d) that it is rated as the most preferred reproduction. We mentioned earlier that stimuli two and six, three and seven, four and eight, and five and nine are essentially redundant pairs and are used to check the observers’ consistency. If the observers are consistent with their selections then we should expect the pattern of stimuli six through nine to resemble that of stimuli two through five. However, the relative scales of stimuli seven and eight do not fully agree with the expectation and a possible explanation is that the subjects especially naïve observers start to notice more details and pay attention to other things such as contrast therefore shifting their priorities as the experiment went along (summarized from the observers’ own descriptions). Regardless of this discrepancy, however, Fig. 12(d) clearly indicates that stimuli two and six are both in the most preferred tier which were corrected using the more preferred sky and green matrices.

As mentioned earlier, the sky variations were dropped for the “Creek” scene and among the stimuli, two and three have the better sky matrix, and two and four have the better green matrix. This scene has different shades of green and rich details, most of the observers described that they had trouble imagining how the actual scene looks, some spent time debating between the images and some just made their decisions intuitively or somewhat randomly. The skin matrices do not produce differences that are noticeable enough so the scales mainly depend on the green matrices showing that the reproduction with the more preferred green matrix is still rated higher than the reproductions with the less preferred green matrix [see Fig. 12(c)]. With the “FlowerPots” scene, some of the pots were detected to have skin colored pixels and the more preferred matrix made them more red whereas the less preferred matrix made them more yellow, most of the observers preferred the less red version of reproduction as we can see in Fig. 12(e) where stimuli four and five have slightly better scores. For the comparison between two and three, and four and five, most observers made their decisions randomly and a small percentage of observers tried hard to distinguish the details in the

background until they found some before their decisions were made. Given enough training and a large enough observer pool, we would expect that stimuli two and three will have very similar scales and so would stimuli four and five. The unconstrained matrix made some of the pots overly red and it fell into the least preferred tier without surprise. The “GreenFlowers” scene [Fig. 12(f)] has rich details in the foreground and some artificial objects in the background. While the stimuli processed using the unconstrained matrix overall are more preferred, the scale differences among them are small, and the chromatic differences are casted by the rich details and contrast variations, which made it difficult for the observers to decide their preferences, most of the decisions were made in a random manner as expected. Similar analysis can be made for Fig. 12(g) and 12(h), for them the observers mostly selected their preference in a coin-toss fashion. Given a large enough observer pool, we would expect the samples to have very similar scales.

For the “Caucasian” scene [Fig. 12(i)], stimulus one is reproduced using the unconstrained matrix (noted as “none” in the plot), which made the skin tone overly red, stimulus two is reproduced using the procedure described in Sec. 4.3 with the more preferred skin matrix (noted by “S”) and the same procedure is used for stimulus three (noted by “s”) with the less preferred skin matrix. Stimulus four is globally corrected using the more preferred skin matrix (noted by “gS”) and stimulus five is globally corrected using the less preferred skin matrix (noted by “gs”). The same manipulations were followed for the “Asian” [Fig. 12(j)] and the “Toys” [Fig. 12(k)] scenes. It is clear from the “Caucasian” and the “Asian” plots that as long as a human face appears in the foreground of the scene and the proportion and details of the face are significant enough to catch the observer’s attention he would give the highest priority to the skin tone before seeking differences from other details. When evaluating the “Toys” scene, it is interesting to see how different sensitivities the observers have: some observers can notice the different color reproductions of the same object with a quick glance, and some did not notice them even at the end of the evaluation. Some of the observers tried hard to recall the colors of the soda cans in the scene and asked the author how the original objects look when debating about their choices. Although there is a good amount of randomness in the evaluation of this scene, the results shown in Fig. 12(k) along with that for the “FlowerPots” scene in Fig. 12(e) can help us conclude that it is probably the best to use the matrix calibrated for the smallest overall color error for scenes consisted of artificial objects.

It is worth noting that we did not make every possible effort for the matrix calibration process to be perfect. When taking the test images and the calibration images, the geometries of the test target or the controls of the camera were not perfectly calibrated. If everything was done in a nearly perfect fashion, the absolute scale differences in Fig. 12 may be different in some of the plots but we expect the same trends to be observed for the stimuli generated using constrained matrices. The nonperfectness is part of the intention of this work whose proposed method does not require perfect calibration conditions while still achieving the desired reproduction.

Another interesting finding is that with the knowledge of a specific observer’s preferences from the rank order experiment, his behavior during the paired comparison evaluation

can be well predicted, indicating that the matrix combination for optimized reproduction can be revealed from the rank order evaluation alone.

5 Conclusions

In this paper, we have presented a hue-constrained matrix optimization method to address the importance of the rotational hue error for preferred color reproduction. By constraining the hue angles of certain patches on a standard test target, we can effectively manipulate the reproduced colors such as skin tone, sky blue, and vegetation green to be subjectively more preferred. We also performed a psychophysical study with carefully selected real-life scenes in order to determine how an optimized color reproduction can be achieved with the proposed method.

For individual memory colors:

- Skin is preferred to have some mild negative angular hue error which yields a healthy looking complexion.
- Sky is preferred to be a pure blue as described by the observers which corresponds to a mild positive hue error. Observers do not like too much green which makes a cyan sky or too much red which results in a purple sky, but cyan is better tolerated than purple.
- Large areas with uniform or nearly uniform colors are more important than areas that are variation and/or texture rich.
- The nature of a memory color determines its position on the priority list when observers evaluate a reproduction with multiple memory colors. Skin is placed on top of the list when a human face is significant enough in the scene to attract the observer's attention. Sky is normally distributed with only small amount of variation in hue or contrast and holds the second place on the list. Grass and vegetation can have many shades of green and can vary in their overall reflectance levels and is given the 3rd place on the list.

For the preferred reproduction of scenes:

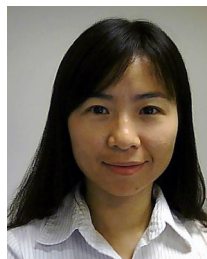
- The preferred matrix for the reproduction of each individual memory color can be found through a rank order experiment with a sufficiently large observer pool.
- For outdoor scenes, a procedure similar to what described in Sec. 4.3 can be used for the preferred overall reproduction and more advanced techniques such as face recognition could be exercised to enhance the color pixel detection accuracy. For indoor scenes, skin tone may be the only memory color to be concerned and a similar approach can be taken as for outdoor scenes.
- For artificial objects, it may be the best to use the matrix calibrated for the smallest overall color error or in other words, the best overall color accuracy

Acknowledgments

We sincerely thank everyone who volunteered during this research for their valuable time. Without their support, understanding and patience, this study would not have been possible.

References

1. G. D. Finlayson and M. S. Drew, "White-point preserving color correction," in *Proc. of the 5th Color Imaging Conference*, IS&T, Albuquerque, New Mexico (1997).
2. P. M. Hubel et al., "Matrix calculations for digital photography," *Proc. of the 5th Color Imaging Conference*, IS&T, Albuquerque, New Mexico (1997).
3. F. W. Billmeyer and M. Saltzman, *Principles of Color Technology*, 3rd ed., John Wiley, New York (2000).
4. C. J. Bartleson, "Color in memory in relation to photographic reproduction," *Photograph. Sci. Eng.* **5**(6), 327–331 (1961).
5. P. Bodrogi and T. Tarczali, "Colourmemory for various sky, skin, and plant colours: effect of the image context," *Color Res. Appl.* **26**(4), 278–289 (2001).
6. K. Töpfer and R. Cookingham, "The quantitative aspects of color quality rendering for memory colors," in *Proc. of PICS*, IS&T, Portland, Oregon (2000).
7. C. J. Bartleson and C. P. Bray, "On the preferred reproduction of flesh, blue-sky, and green-grass colors," *Photogr. Sci. Eng.* **6**(1), 19–25 (1962).
8. C. J. Bartleson, "Memory colors of familiar objects," *J. Opt. Soc. A* **50**(1), 73–77 (1960).
9. S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Color reproduction and the naturalness constraint," *Color Res. Appl.* **24**(4), 52–67 (1999).
10. CIE, *Guidelines for the evaluation of Gamut Mapping Algorithms* (2004).
11. S. R. Fernandez and M. D. Fairchild, "Preferred color reproduction of images with unknown colorimetry," in *9th Color Imaging Conference*, IS&T, San Jose, California (2001).
12. Y. Kuwahara et al., "Image processing device, image processing system, image processing method, image processing program, and integrated circuit device," US patent 8,000,525. (2011).
13. J. D. Buhr and K. Topfer, "Color image reproduction of scenes with preferential color mapping," US patent 6,791,716 (2004).
14. K. Hara et al., "Preferred color reproduction based on personal histogram transformation," *IEEE Tran. Consumer Electron.* **55**(2), 855–863 (2009).
15. S. Kang, O. Kwon, and S. Chien, "Preferred skin color reproduction based on y-dependent gaussian modeling of skin color," *J. Imag. Sci. Technol.* **55**(4), 1–7 (2011).
16. S. Quan and E. Jin, "Memory color based preferred color reproduction with psychophysical evaluation," in *Proc. Color Imaging Conference - Proc. of the 16th Color Imaging Conference*, IS&T, Portland, Oregon (2008).
17. G. D. Finlayson and M. S. Drew, "Constrained least-squares regression in color spaces," *J. Electron. Imag.* **6**(4), 484–493 (1997).
18. S. Quan, "Analytical approach to the optimal linear matrix with comprehensive error metric," *Proc. SPIE* **5292**, 243–253 (2004).
19. G. Sharma, *Digital Color Imaging Handbook*, CRC Press, Florida (2003).
20. W. Hock and K. Schittkowski, "A comparative performance evaluation of 27 nonlinear programming coded," *Computing* **30**(4), 335–358 (1983).
21. K. Schittkowski, "NLPQL: a FORTRAN-subroutine solving constrained nonlinear programming problems," *Ann. Operat. Res.* **5**(1–4), 485–500 (1986).
22. R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.* **6**(2), 163–168 (1963).
23. M. H. Wright, "The interior-point revolution in optimization: history, recent developments, and lasting consequences," *Bull. Amer. Math. Soc.* **42**(1), 39–56 (2005).
24. T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optimiz.* **6**(2), 418–445 (1996).
25. H. Zhang, H. Liu, and S. Quan, "Color correction optimization with hue regularization," *Proc. SPIE* **7866**, 786610 (2011).
26. X. Zhang and R. Bayer, "Perceptual preference for noise and color saturation tradeoff in digital camera images," in *Proc. Color Imaging Conference*, IS&T, Scottsdale, AZ (2006).
27. H. Zhang and S. Quan, "Memory color assisted illuminant estimation through pixel clustering," *Proc. SPIE* **7537**, 75370J (2010).



Heng Zhang received her BS degree from Tsinghua University, Beijing, China and her MS degree from Oregon State University, both in electrical engineering. She is currently pursuing a PhD degree with Oregon State University in electrical engineering with a focus in image processing. She currently works as a firmware engineer at Broadcom Corporation developing ISP drivers and tools for cellular applications. She previously worked for Micron Technology (now Aptina Imaging) as an imaging engineer and ESS Technology as an algorithm/firmware engineer.



Huaping Liu received his BS and MS degrees in electrical engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 1987 and 1990, respectively, and his PhD degree in electrical engineering from New Jersey Institute of Technology, Newark, in 1997. From 1997 to 2001, he was with Lucent Technologies, Whippany, NJ. He joined the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, in 2001,

and has been a professor since 2011. His research interests include signal processing, ultrawideband systems, multiple-input multiple-output antenna systems, channel coding, and modulation and detection techniques for multiuser communications. He has served as an associate editor for the *IEEE Transactions on Vehicular Technology* and *IEEE*

Communications Letters from 2009 to 2011. He is currently an editor for the *Journal of Communications and Networks*.



Shuxue Quan is currently senior principal scientist at Broadcom Corporation. He received his PhD in imaging science from Rochester Institute of Technology in 2002. He has been working at Sony, Qualcomm, Micron/Aptina and Broadcom for several years each on various areas. He is interested in the research and development on imaging technologies, mobile multimedia and application processors, and computer vision. He has over 10 granted patents and published more than 20 conference and journal papers.