

Real-Time Spectral Rendering

C. Allen Waddle; University of Victoria; Victoria, British Columbia, Canada

Abstract

The common practice in computer graphics of multiplying RGB triplets to model the perception of reflected light is simple and efficient, but it does not predict color reliably, nor can it model metamerism or optical phenomena that are spectral in nature, such as iridescence. Spectral rendering meets these goals, but the $O(n)$ cost incurred in multiplication of n -dimensional spectra can be prohibitively expensive. Although spectra can be well approximated by linear combinations of $m \ll n$ basis vectors, standard linear models lead to matrix-vector multiplication with complexity $O(m^2)$ when performing lighting calculations in the lower-dimensional space. A method by Drew and Finlayson reduces this cost to $O(m)$ with a “sharp” basis that has been transformed by an $m \times m$ matrix \mathbf{T} , each column of which is the solution to an optimization problem requiring specification of an interval of wavelengths. Choosing good intervals, however, is itself an optimization problem, which the authors neither pose nor solve. Instead, we construct \mathbf{T} by optimizing the sharpness of a basis and the minimum angle between its vectors, but obtain better results by minimizing the residual error explicitly. Alternatively, by minimizing a weighted subspace projection we can solve a simpler problem that converges to the same minimum-residual solutions. Testing these methods with a variety of spectra, we obtain good accuracy with as few as four dimensions, permitting real-time rendering at arbitrarily high spectral resolutions for a cost that is only a fraction above that of RGB rendering.

Introduction

A color signal arriving at the eye is characterized by a *spectral power distribution* (SPD) of the radiant power present at each wavelength within the visual range. Most of this light arrives from a source, or *illuminant*, after interaction with a material that absorbs, reflects and/or transmits different wavelengths of incident light according to the optical properties that emerge from its molecular structure. Although the methods described here can model transmittance, we confine our discussion to Lambertian surface reflectance by normal materials (e.g., non-fluorescent), which is fully characterized by a *reflectance spectrum* giving the fraction of incident light reflected at each wavelength. Furthermore, we use the word *spectrum* to refer to an SPD or a reflectance spectrum interchangeably.

Because human color vision is three-dimensional, a triplet representing relative proportions of primary colors is an appropriate description of a color, which is the subjective *result* of light interacting with a material and an observer’s visual system. Common practice in computer graphics omits the explicit notion of an observer and uses RGB triplets in a different way to describe the spectral *causes* of color. That is, spectra are represented by coarse samplings at red, green and blue wavelengths, and their interaction is modeled by componentwise multiplication, which amounts to multiplying the color of an illuminant by the color of a surface illuminated by a white light. As shown by Borges [3], this efficient approximation fails to predict color when spectra differ sufficiently from constant values.

It also fails to model metamerism, as the colors of a pair of surfaces represented as triplets will always either match or not match under any illuminant with non-zero RGB components. Nor can it model interference effects, such as iridescence, which require information that is missing from the RGB model.

With spectra represented as n -dimensional vectors of point samples, color signals can be formed with complexity $O(n)$ by componentwise multiplication. Display colors are then computed by a linear mapping to coordinates in a three-dimensional color space. Finite-dimensional linear models may be used to represent spectra as vectors of $m \ll n$ coefficients multiplying basis vectors, but the computational benefits of this method are lost when $m^2 \geq n$, because color signals are computed with matrix-vector multiplication, with complexity $O(m^2)$.

To reduce this cost, Drew and Finlayson propose a method in which basis vectors are “sharpened” by a change of coordinates [7]. If the transformed vectors are sufficiently disjoint, their coefficients may be multiplied with acceptable error when the transformation is inverted. To find the change-of-basis matrix \mathbf{T} , m optimization problems are solved, each of which requires the choice of an interval of wavelengths in which the corresponding basis vector is sharpened. Finding good intervals, however, is itself an optimization problem, which the authors do not address. Having found that good choices are not obvious, we instead pose and solve a nonlinear programming (NLP) problem to find the matrix \mathbf{T} that minimizes the residual error. With this method we obtain accurate results with four to nine dimensions when rendering scenes involving arbitrary combinations of daylight, incandescent and fluorescent illumination, with performance comparable to that of RGB rendering.

Because explicit minimization of the residual error is impractical in many applications, which may involve millions of different reflectance spectra, an appealing alternative is to optimize a simpler function of the sharp basis. An intuitive choice is to maximize sharpness, but we find that the sharpest basis does not yield the smallest residual error, even when the basis vectors are close to orthogonal. Instead, by minimizing a weighted projection of the subspace spanned by the reflectance spectra onto a subspace that is a function of the illuminant(s) and \mathbf{T} we obtain results identical to those obtained by minimizing the residual error explicitly. The complexity of this simpler optimization problem is independent of the number of reflectance spectra and depends only on a small number of vectors, obtained by a singular value decomposition (SVD) of the reflectance spectra of interest, which together form an orthonormal basis for the subspace spanned by the whole set.

The RGB Approximation

A color signal $C(\lambda) = E(\lambda)S(\lambda)$ is the product of an SPD $E(\lambda)$ and a reflectance spectrum $S(\lambda)$, both continuous functions of wavelength λ . The formation of a color \mathbf{b} from an observer’s response is

approximated in discrete form by a matrix-vector multiplication¹:

$$\mathbf{b} = \left(\int R(\lambda)C(\lambda)d\lambda, \int G(\lambda)C(\lambda)d\lambda, \int B(\lambda)C(\lambda)d\lambda \right)^T \\ \approx \mathbf{M}_{XYZ \rightarrow RGB} \mathbf{M}_{SPD \rightarrow XYZ} (\mathbf{E} * \mathbf{S}) = \mathbf{MC} = (r, g, b)_C^T,$$

where $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$ are an observer's trichromatic sensitivities, \mathbf{E} , \mathbf{S} and \mathbf{C} are n -dimensional vectors of point samples of the spectra $E(\lambda)$, $S(\lambda)$, and $C(\lambda)$, and the $3 \times n$ matrix \mathbf{M} is the product of a $3 \times n$ matrix containing the International Commission on Illumination (CIE) 1931 2° Standard Observer color-matching functions and a 3×3 XYZ-to-RGB conversion matrix [1]. We use 61-dimensional vectors to represent spectra sampled between 400 and 700 nm at 5 nm intervals, a resolution high enough to capture the narrow peaks of typical fluorescent illuminants [15].

The usual method of rendering with RGB triplets multiplies the color of an illuminant by the color of a reflectance spectrum illuminated with equal energy at unity amplitude:

$$\mathbf{b}' = \mathbf{ME} * \mathbf{MS} = (r, g, b)_E^T * (r, g, b)_S^T. \quad (1)$$

In general, $\mathbf{b}' \neq \mathbf{b}$, but this approximation is very good in practice, if either spectrum is close enough to a scaled white [3]. In these cases Eq. (1) is as much a color *specification* as it is a *prediction* of the color that would be observed by a human with normal color vision when light from illuminant \mathbf{E} is reflected by the surface \mathbf{S} .

Finite-Dimensional Linear Models of Spectra

Typical spectra have been shown to be well approximated by linear combinations of a small number of orthonormal basis vectors obtained by principal component analysis (PCA) [17]. We represent SPDs and reflectance spectra with a single basis, obtained from a PCA of the color signals formed by p illuminant SPDs and q reflectance spectra, retaining the first m principal components as the columns of an $n \times m$ orthonormal matrix \mathbf{Q} . Coefficients representing spectra \mathbf{S} and \mathbf{E} in the m -dimensional basis are thus $\mathbf{s} = \mathbf{Q}^T \mathbf{S}$ and $\mathbf{e} = \mathbf{Q}^T \mathbf{E}$.

Computing a color signal in the m -dimensional space involves an implicit expansion into n dimensions: $\mathbf{c} = \mathbf{Q}^T \text{diag}(\mathbf{S}) \mathbf{Q} \mathbf{e} = \mathbf{R}_s \mathbf{e}$, where the $m \times m$ matrix \mathbf{R}_s is a function of the reflectance spectrum \mathbf{S} . Thus we see that rendering in m dimensions requires matrix-vector multiplication with $O(m^2)$ cost. An approximate color signal $\mathbf{C} \approx \mathbf{Q} \mathbf{c}$ is obtained by a weighted combination of the basis vectors.

Spectral Sharpening

To reduce the cost of rendering in a lower-dimensional space, Drew and Finlayson propose the method of spectral sharpening [7], in which a linear transformation by an $m \times m$ matrix \mathbf{T} forms a new set of ideally disjoint, or *complementary*, basis vectors $\tilde{\mathbf{Q}} = \mathbf{Q} \mathbf{T}$. That is, if distinct basis vectors $\tilde{\mathbf{Q}}_i$ and $\tilde{\mathbf{Q}}_j$ are complementary, then $\tilde{\mathbf{Q}}_i * \tilde{\mathbf{Q}}_j = \mathbf{0}$. If this condition is met, the computation of an exact color signal in the sharp basis is reduced to componentwise multiplication of vectors of coefficients representing an illuminant and a reflectance, as in the RGB approximation, but in m dimensions instead

¹We denote matrices and vectors with boldface or calligraphic letters, and all vectors are assumed to be column vectors, unless otherwise noted. An asterisk is used to denote componentwise vector multiplication; that is, for vectors $\mathbf{U} = (U_1, U_2, \dots, U_n)^T$ and $\mathbf{V} = (V_1, V_2, \dots, V_n)^T$, $\mathbf{U} * \mathbf{V} = \mathbf{V} * \mathbf{U} = \text{diag}(\mathbf{U}) \mathbf{V} = \text{diag}(\mathbf{V}) \mathbf{U} = (U_1 V_1, U_2 V_2, \dots, U_n V_n)^T$.

of three. For nonsingular \mathbf{T} , no error is incurred by the sharpening transformation itself, because it is invertible. Rather, error results from multiplication of sharp basis coefficients, since in general the sharp basis vectors are only approximately complementary.

Because $\tilde{\mathbf{Q}}$ is not an orthonormal basis, computing sharp basis coefficients involves the Moore-Penrose pseudoinverse $\tilde{\mathbf{Q}}^+ = (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T = \mathbf{T}^{-1} \mathbf{Q}^T$, so that $\tilde{\mathbf{s}} = \tilde{\mathbf{Q}}^+ \mathbf{S}$ and $\tilde{\mathbf{e}} = \tilde{\mathbf{Q}}^+ \mathbf{E}$. An approximate color signal is thus computed as

$$\mathbf{C} \approx \tilde{\mathbf{Q}}(\tilde{\mathbf{e}} * \tilde{\mathbf{s}}) = \mathbf{Q} \mathbf{T} [(\mathbf{T}^{-1} \mathbf{Q}^T \mathbf{E}) * (\mathbf{T}^{-1} \mathbf{Q}^T \mathbf{S})] \\ = \mathbf{Q} \mathbf{T} [(\mathbf{T}^{-1} \mathbf{e}) * (\mathbf{T}^{-1} \mathbf{s})]. \quad (2)$$

Eq. (2) shows that rendering with sharp basis coefficients amounts to changing the basis in which coefficients are multiplied. The product is then transformed back to the standard basis before expansion. Resulting color signals can contain negative components, which we clamp to zero. For multiple reflections, sharp basis coefficients are multiplied sequentially, $\mathbf{C} \approx \tilde{\mathbf{Q}} \tilde{\mathbf{s}}_k * \dots * \tilde{\mathbf{s}}_2 * \tilde{\mathbf{s}}_1 * \tilde{\mathbf{e}}$.

In the ideal case, for columns $\tilde{\mathbf{Q}}_i$ and $\tilde{\mathbf{Q}}_j$ of $\tilde{\mathbf{Q}}$, we have

$$\mathbf{C} = \tilde{\mathbf{Q}} \tilde{\mathbf{e}} * \tilde{\mathbf{Q}} \tilde{\mathbf{s}} = \sum_{i=1}^m \tilde{\mathbf{Q}}_i * \tilde{\mathbf{Q}}_i \tilde{e}_i \tilde{s}_i.$$

To form \mathbf{C} with linear combinations of $\tilde{\mathbf{Q}}_i$ rather than $\tilde{\mathbf{Q}}_i * \tilde{\mathbf{Q}}_i$, sharp basis coordinates are defined as relative to the coordinates $\tilde{\mathbf{w}}$ of a white reflectance $\mathbf{W} = \mathbf{1}$. Thus,

$$\frac{\mathbf{C}}{\mathbf{W}} = \sum_{i=1}^m \frac{\tilde{\mathbf{Q}}_i * \tilde{\mathbf{Q}}_i \tilde{e}_i \tilde{s}_i}{\tilde{\mathbf{Q}}_i \tilde{w}_i} = \sum_{i=1}^m \tilde{\mathbf{Q}}_i \frac{\tilde{e}_i \tilde{s}_i}{\tilde{w}_i} = \sum_{i=1}^m \tilde{\mathbf{Q}}_i \tilde{e}_i \tilde{s}_i.$$

The last equality is satisfied if $\tilde{\mathbf{w}} = \mathbf{1}$, which is accomplished by normalizing $\tilde{\mathbf{Q}}^+$ so that $\tilde{\mathbf{Q}}^+ \mathbf{1} = \mathbf{1}$. Equivalently, we may normalize \mathbf{T} , so that $\mathbf{T} \mathbf{1} = \mathbf{Q}^T \mathbf{1}$. In our optimizations that minimize the residual error, or an equivalent weighted subspace projection, this constraint is not strictly necessary, as the optimal solutions all converge to solutions with this property. However, some optimization algorithms may benefit from imposing it explicitly.

It should be noted that as a consequence of defining “color” in the sharp basis as relative to white, color signals formed from a scaled white light or surface are independent of \mathbf{T} and exactly equal to their projections onto the orthonormal basis \mathbf{Q} :

$$\tilde{\mathbf{Q}}[(\tilde{\mathbf{Q}}^+ \alpha \mathbf{W}) * (\tilde{\mathbf{Q}}^+ \mathbf{S})] = \alpha \tilde{\mathbf{Q}} \tilde{\mathbf{Q}}^+ \mathbf{S} = \alpha \mathbf{Q} \mathbf{Q}^T \mathbf{S} = \alpha \mathbf{Q} \mathbf{s} \\ \tilde{\mathbf{Q}}[(\tilde{\mathbf{Q}}^+ \mathbf{E}) * (\tilde{\mathbf{Q}}^+ \alpha \mathbf{W})] = \alpha \tilde{\mathbf{Q}} \tilde{\mathbf{Q}}^+ \mathbf{E} = \alpha \mathbf{Q} \mathbf{Q}^T \mathbf{E} = \alpha \mathbf{Q} \mathbf{e}.$$

In these cases the error in the sharp basis approximation to a finite-dimensional linear model vanishes.

In [7] the authors switch the roles of $\tilde{\mathbf{Q}}^T$ and $\tilde{\mathbf{Q}}^+$, computing sharp basis coefficients as $\tilde{\mathbf{s}} = \tilde{\mathbf{Q}}^T \mathbf{S}$ and $\tilde{\mathbf{e}} = \tilde{\mathbf{Q}}^T \mathbf{E}$ and color signals as $\mathbf{C} = (\tilde{\mathbf{Q}}^+)^T \tilde{\mathbf{e}} * \tilde{\mathbf{s}}$. Since $(\tilde{\mathbf{Q}}^+)^+ = \tilde{\mathbf{Q}}$, this notation is equivalent to ours, which we prefer if only because it is standard and familiar. It should be noted, however, that their matrix \mathbf{T} is therefore equal to the transpose of the inverse of ours, since $(\tilde{\mathbf{Q}}^+)^T = \mathbf{Q} \mathbf{T}^{-T}$. This is also an equivalent notation, as specification of \mathbf{T} determines \mathbf{T}^{-T} and vice versa. On these two points our formulations are equivalent, but an important difference may occur in the method used to find \mathbf{T} . Drew and Finlayson maximize the sharpness of $\tilde{\mathbf{Q}}^+$, while we maximize the sharpness of $\tilde{\mathbf{Q}}$. Since, as we show below, the latter

approach does indeed yield better results, we suspect that the effectiveness of their method is due to the fact that sharpening of $\tilde{\mathbf{Q}}^+$ tends to sharpen $\tilde{\mathbf{Q}}$ implicitly, if not maximally. That this is the case can be seen by noting that since $\tilde{\mathbf{Q}}^+ \tilde{\mathbf{Q}} = \mathbf{I}$, for rows i and columns j we have $\langle \tilde{\mathbf{Q}}_i^+, \tilde{\mathbf{Q}}_j \rangle = \delta_{ij}$. Thus, constraining the rows of $\tilde{\mathbf{Q}}^+$ to be approximately complementary will tend to do the same for the columns of $\tilde{\mathbf{Q}}$.

Finding \mathbf{T}

The effectiveness of spectral sharpening depends on finding a matrix \mathbf{T} that transforms \mathbf{Q} in such a way that the error incurred by multiplication of sharp basis coefficients is reduced to an acceptable level. In [7] this is posed as an optimization problem requiring the specification of a set of m wavelength intervals, which the authors simply suggest should be approximately uniform and evenly spaced over the visual range. Metaheuristic methods, such as particle swarm optimization (PSO) [11], have shown us that good intervals are in general neither uniform nor evenly spaced.

Removing the need for sharpening intervals, we obtain better results by solving a nonlinear optimization problem that seeks to minimize the residual error. Alternatively, maximizing the sharpness of the transformed basis should be considerably less expensive, as it involves only $\binom{m}{2}$ pairs of vectors instead of $p \cdot q$ color signals. We find, however, that a sharper basis does not in general yield a lower residual error. Hypothesizing that the quality of a sharp basis also depends on the angles between the basis vectors, we pose an optimization problem that maximizes sharpness while constraining the minimum angle over a range of angles and choose from among the optimal \mathbf{T} matrices the one that yields the smallest residual error. Although we do find that the bases obtained by this approach are better than those found by simply maximizing sharpness, they are not in general as good as those found by minimizing the residual error. Instead, seeking an alternative approach whose complexity is also independent of the size of a color signal set, we pose a problem that minimizes the weighted projection of a d -dimensional subspace spanned by the reflectance spectra, where $d < n \ll q$, onto a subspace spanned by a function of \mathbf{T} and the illuminants. The sharp bases found by this more efficient method are equivalent to those found by minimizing the residual error explicitly.

By Evenly Spaced Sharpening Intervals

Drew and Finlayson construct the m columns of \mathbf{T}^{-T} by sharpening the corresponding rows of $\tilde{\mathbf{Q}}^+$ one at a time within evenly spaced wavelength intervals. In [7] this is posed as m quadratic programming problems with a non-negativity constraint on $\tilde{\mathbf{Q}}^+$. We adopt instead a method described by the same authors in an earlier article on spectral sharpening [9], in which each column of \mathbf{T}^{-T} is the solution to an eigenvector problem which allows the sharp basis to take on negative values. By removing non-negativity constraints on $\tilde{\mathbf{Q}}^+$ and $\tilde{\mathbf{Q}}$, which are not mathematically necessary, from this and all other methods described here, we would expect to obtain sharp bases that yield smaller residual errors.

By Optimizing the Sharpening Intervals

With $\binom{n}{2m}$ possible choices, an exhaustive search for the best sharpening intervals is not practical. Instead, we use PSO to find

intervals that minimize the total squared residual error:

$$\begin{aligned} & \sum_{a=1}^p \sum_{b=1}^q \|\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_a * \tilde{\mathbf{s}}_b - \tilde{\mathbf{Q}}^+ \mathbf{E}_a * \mathbf{S}_b)\|_2^2 \\ &= \sum_{a=1}^p \sum_{b=1}^q \|\tilde{\mathbf{Q}}[\text{diag}(\tilde{\mathbf{e}}_a) \tilde{\mathbf{Q}}^+ - \tilde{\mathbf{Q}}^+ \text{diag}(\mathbf{E}_a)] \mathbf{S}_b\|_2^2 \\ &= \sum_{a=1}^p \|\mathbf{T}[\text{diag}(\tilde{\mathbf{e}}_a) \tilde{\mathbf{Q}}^+ - \tilde{\mathbf{Q}}^+ \text{diag}(\mathbf{E}_a)] \mathbf{S}\|_F^2 \\ &= \sum_{a=1}^p \|\Delta \tilde{\mathbf{e}}_a \mathbf{S}\|_F^2 = \|\Delta \tilde{\mathbf{e}}_1 \cdots \Delta \tilde{\mathbf{e}}_p\|^T \mathbf{S}\|_F^2 = \|\Delta \tilde{\mathbf{e}}^T \mathbf{S}\|_F^2, \end{aligned} \quad (3)$$

where the $n \times p \cdot m$ matrix $\Delta \tilde{\mathbf{e}}$ is a function of \mathbf{T} and p illuminants, \mathbf{S} is a $n \times q$ matrix containing q reflectance spectra and $\|\cdot\|_F$ is the Frobenius matrix norm.

Each particle visits a sequence of points along its trajectory in $2m$ -dimensional space, with the coordinates of every point $\mathbf{x} = (x_1, \dots, x_{2m})$ constrained so that $x_i \in \{400, 405, \dots, 700\}$ and $x_i < x_{i+1}$. Each point corresponds to a set of sharpening intervals that yields the m columns of \mathbf{T}^{-T} by the method described in [9]. The best set of visited intervals is the one that produces the sharp basis $\tilde{\mathbf{Q}}$ yielding the smallest residual error as expressed by Eq. (3).

By Minimizing the Residual Error

Rather than search for the sharpening intervals that minimize the residual error, we may minimize the error explicitly by solving

$$\min. \|\Delta \tilde{\mathbf{e}}^T \mathbf{S}\|_F^2 \quad (4)$$

with an NLP method, such as sequential quadratic programming or the Levenberg-Marquardt algorithm.

By Optimizing Sharpness and the Minimum Angle

In [6] the authors suggest that the quality of a sharp basis is due to the sharpness of the pseudoinverse and the angles between its rows. Angles close to 90° are preferred, because they reduce ‘‘cross talk.’’ Sharpness and angles, however, are inseparable features. Nevertheless, without knowledge of the function that relates them to the residual error, we can try to optimize their combination by maximizing sharpness while constraining the minimum angle between any pair of vectors to a sequence of angles. Angles constrained to a minimum of 0° are effectively unconstrained. From among the resulting \mathbf{T} matrices we choose the one that yields the least residual error.

$$\min._{\theta} \|\Delta \tilde{\mathbf{e}}^T \mathbf{S}\|_F^2 \quad (5a)$$

$$\text{s.t.} \quad \min_{\mathbf{T}} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \frac{\langle \tilde{\mathbf{Q}}_i^+ * \tilde{\mathbf{Q}}_i^+, \tilde{\mathbf{Q}}_j^+ * \tilde{\mathbf{Q}}_j^+ \rangle}{\langle \tilde{\mathbf{Q}}_i^+, \tilde{\mathbf{Q}}_i^+ \rangle \langle \tilde{\mathbf{Q}}_j^+, \tilde{\mathbf{Q}}_j^+ \rangle} \quad (5b)$$

$$\frac{\langle \tilde{\mathbf{Q}}_i^+, \tilde{\mathbf{Q}}_j^+ \rangle^2}{\langle \tilde{\mathbf{Q}}_i^+, \tilde{\mathbf{Q}}_i^+ \rangle \langle \tilde{\mathbf{Q}}_j^+, \tilde{\mathbf{Q}}_j^+ \rangle} \leq \cos^2(\theta), \quad \theta = 90^\circ, \dots, 0^\circ \quad (5c)$$

$$\mathbf{T} \mathbf{1} = \mathbf{Q}^T \mathbf{1}. \quad (5d)$$

By substituting columns $\tilde{\mathbf{Q}}_i$ and $\tilde{\mathbf{Q}}_j$ of $\tilde{\mathbf{Q}}$ for rows $\tilde{\mathbf{Q}}_i^+$ and $\tilde{\mathbf{Q}}_j^+$ of $\tilde{\mathbf{Q}}^+$ in Eqs. (5b) and/or (5c), we obtain four NLP problems. Doing so allows us to test the hypotheses that the sharpness of $\tilde{\mathbf{Q}}$ is in fact more important than that of $\tilde{\mathbf{Q}}^+$, and that both angle constraints yield equivalent results.

By Minimizing a Weighted Subspace Projection

Observing that residual error is minimized by minimizing the projection of the column space of \mathcal{S} onto the column space of $\Delta\tilde{\mathbf{E}}$, we may find \mathbf{T} by forming the projection matrix $\Delta\tilde{\mathbf{E}}\Delta\tilde{\mathbf{E}}^+$ and solving

$$\min. \left\| \Delta\tilde{\mathbf{E}} \Delta\tilde{\mathbf{E}}^+ \mathcal{S} \right\|_F^2.$$

We can instead solve a smaller problem by constructing an orthonormal basis for the column space of \mathcal{S} from its SVD, $\mathcal{S} = \mathbf{U}_S \mathbf{\Sigma}_S \mathbf{V}_S^T$, where \mathbf{U}_S and \mathbf{V}_S are respectively $n \times n$ and $q \times n$ orthonormal bases for the column and row spaces of \mathcal{S} , and $\mathbf{\Sigma}_S$ is an $n \times n$ diagonal matrix of its singular values arranged in descending order. Because a small number of dimensions accounts for almost all the variance in spectra, in the typical case only the first $d < n$ columns of \mathbf{U}_S are of interest, which we weight by their corresponding singular values. Likewise, the decomposition $\Delta\tilde{\mathbf{E}} = \mathbf{U}_{\Delta\tilde{\mathbf{E}}} \mathbf{\Sigma}_{\Delta\tilde{\mathbf{E}}} \mathbf{V}_{\Delta\tilde{\mathbf{E}}}^T$ yields $\mathbf{U}_{\Delta\tilde{\mathbf{E}}} \mathbf{\Sigma}_{\Delta\tilde{\mathbf{E}}}$, a weighted basis for the column space of $\Delta\tilde{\mathbf{E}}$. Our smaller NLP problem is thus

$$\min. \left\| \mathbf{\Sigma}_{\Delta\tilde{\mathbf{E}}} \mathbf{U}_{\Delta\tilde{\mathbf{E}}}^T \mathbf{U}'_S \mathbf{\Sigma}'_S \right\|_F^2, \quad (6)$$

where \mathbf{U}'_S is the first d columns of \mathbf{U}_S , and diagonal $\mathbf{\Sigma}'_S$ contains the corresponding singular values. Since \mathbf{U}'_S and $\mathbf{\Sigma}'_S$ are not functions of \mathbf{T} , they are computed only once, before an optimization begins.

Real-Time Implementation

Our method's computations are performed during a one-time pre-process or once every frame in a real-time application. As most of these fall in the first category, the real-time computational complexity is similar to that of RGB rendering. Once a sharp basis is constructed and corresponding coefficients are pre-computed for all spectra, rendering consists simply of lighting calculations performed by multiplying sharp basis coefficients.

When all of a frame's lighting calculations are completed, the linear RGB color at each pixel is computed by multiplying a vector of sharp basis coefficients by a $3 \times m$ matrix $\mathbf{M}\mathbf{Q}$, where the 3×3 matrix \mathbf{M} is the concatenation of: 1) a linear mapping of SPD to XYZ coordinates; 2) a chromatic adaptation; and 3) a linear mapping from adapted XYZ to RGB coordinates. To compute the chromatic adaptation, various methods exist for estimating a scene's currently dominant illumination, such as the gray world assumption [4], or white patch algorithms, including the Retinex method [14]. Because the assumptions on which these methods are based can be easily violated, we prefer the method of Wilkie and Weidlich [20], which is general and robust.

Results

To test our methods we use a representative from each of three classes of illuminants: D65 for daylight, illuminant A for incandescent light and the fluorescent F2. For reflectance spectra, we use three sets for their different characteristics. The 1,269 Munsell Matte spectra measured by the University of Eastern Finland Spectral Color Research Group² [12] are familiar to the color science community. As well, they are generally smooth enough to be well approximated in a small number of dimensions. We also use the same group's Natural Colors, a set of 218 spectra of mostly colorful plant materials. These more challenging spectra, many of which

²<http://www.uef.fi/fi/spectral/spectral-database>

yield saturated colors, generally require more dimensions to approximate with similar accuracy. Finally, we include the 48 spectra of Fairchild's and Johnson's METACOW image [8], both because half of them are measured from the well known ColorChecker color rendition chart³ [18] and because the other half are particularly challenging. These synthetic spectra are constructed by adding linear combinations of metameric blacks to yield colors that match when illuminated by D65 and differ significantly when illuminated by illuminant A. Being far from smooth, these unnatural spectra are an excellent stress test.

Accuracy

For each set of reflectance spectra we use the three illuminants to construct four color signal sets. Three of these consist of the reflectance spectra illuminated by a single illuminant; the fourth consists of their union. For each of these sets we use PCA to construct six m -dimensional orthonormal bases, where $m \in \{4, 5, \dots, 9\}$. For each orthonormal basis \mathbf{Q} we use each of our methods to find a matrix \mathbf{T} . Accuracy is measured by comparing the color signals obtained by componentwise multiplication of n -dimensional spectra, considered to be ground truth, to those obtained by componentwise multiplication of sharp basis coefficients. For sets constructed with a single illuminant, we thus measure residual errors obtained from approximations of $q \in \{1269, 218, 48\}$ color signals. For sets constructed with three illuminants, we measure four times these numbers by also comparing color signals constructed with an illuminant that is the average of all three.

We use a normalized root-mean-square error (NRMSE) to measure residual spectral error:

$$\text{NRMSE} = \frac{100\%}{\max_{i,\lambda} C_i(\lambda)} \sqrt{\frac{\sum_{\lambda} [(\tilde{\mathbf{Q}}\tilde{\mathbf{e}} * \tilde{\mathbf{s}})(\lambda) - \mathbf{C}(\lambda)]^2}{n}},$$

where $i \in \{1, \dots, p \cdot q\}$ indexes the color signals in a set. That is, to set a convenient and intuitive scale, the RMS error is normalized as a percentage of the maximum amplitude.

We measure the effect of spectral error on color accuracy with the CIEDE2000 color difference formula [5]. A commonly cited rule of thumb is that a difference of 1 is just noticeable, but differences as high as 3 or more may be imperceptible [16]. In any case, color difference measurements should not be interpreted as absolute measures of accuracy, because the viewing conditions for which they are defined are quite different from ours. In particular, the CIEDE2000 formula measures color differences of large patches on a uniform gray background with D65 illumination. Nevertheless, pixel color difference may be a useful measure of perceptibility of differences between images, and in at least one study it was found that differences between pairs of images with average pixel differences of $\Delta E_{ab} < 2.15$ were imperceptible [19]. Thus, caveats notwithstanding, the CIEDE2000 color difference measurements included here should provide an approximate, relative scaling of the perceptual effects of spectral errors.

The results depicted in Fig. 1, showing average spectral errors and color differences obtained for single-illuminant METACOW color signal sets, are representative of the results obtained for all other single- and multi-illuminant sets when sharpness is optimized. For all D65 sets, optimizing the sharpness of either $\tilde{\mathbf{Q}}$ or $\tilde{\mathbf{Q}}^+$ yields

³ColorChecker Classic at xritephoto.com.

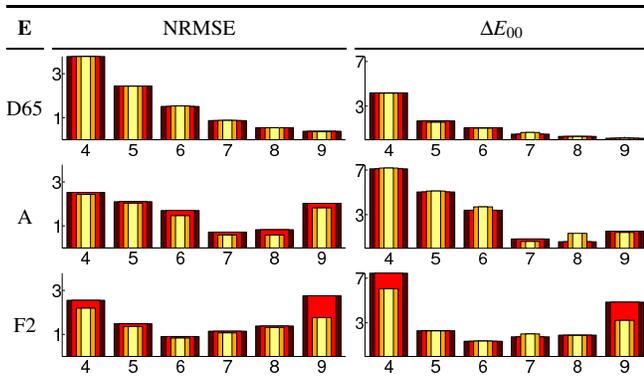


Figure 1. For dimensions $m = 4, 5, \dots, 9$, average spectral errors and color differences of METACOW color signals computed with single-illuminant, sharp bases constructed by methods that optimize: from outer (darker) to inner (lighter) bars 1) \tilde{Q}^+ sharpness and its minimum angle; 2) \tilde{Q}^+ sharpness and the minimum \tilde{Q} angle; 3) \tilde{Q} sharpness and the minimum \tilde{Q}^+ angle; and 4) \tilde{Q} sharpness and its minimum angle.

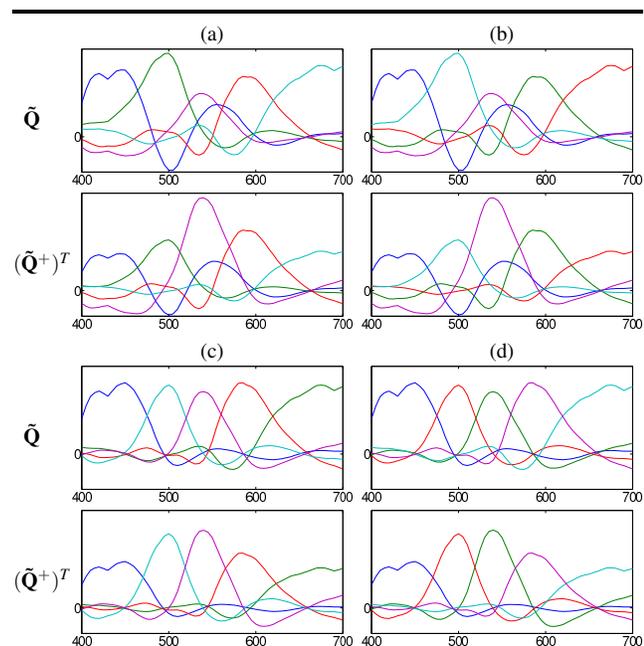


Figure 2. 5-dimensional sharp bases and their pseudoinverses obtained from Munsell/D65 color signals by: a) minimizing the residual error explicitly; b) minimizing a weighted subspace projection; c) optimizing the sharpness of \tilde{Q} and its minimum angle; d) optimizing the sharpness of \tilde{Q} the minimum \tilde{Q}^+ angle.

similar results, but significant differences are seen with F2 and illuminant A. The best results, in general, are obtained by maximizing the sharpness of \tilde{Q} while constraining the minimum angle between its columns or between the rows of \tilde{Q}^+ . For all color signal sets and all dimensions, optimization with either of these constraints yields an essentially identical result, as shown, for Munsell/D65 color signals and $m = 5$, by the bottom two rows of Fig. 2. Perhaps the most significant shortcoming of the methods that optimize sharpness is that their accuracy is not necessarily improved by simply increasing the dimensionality of a basis, which, for METACOW color signals formed with F2 and illuminant A, is shown clearly in Fig. 1.

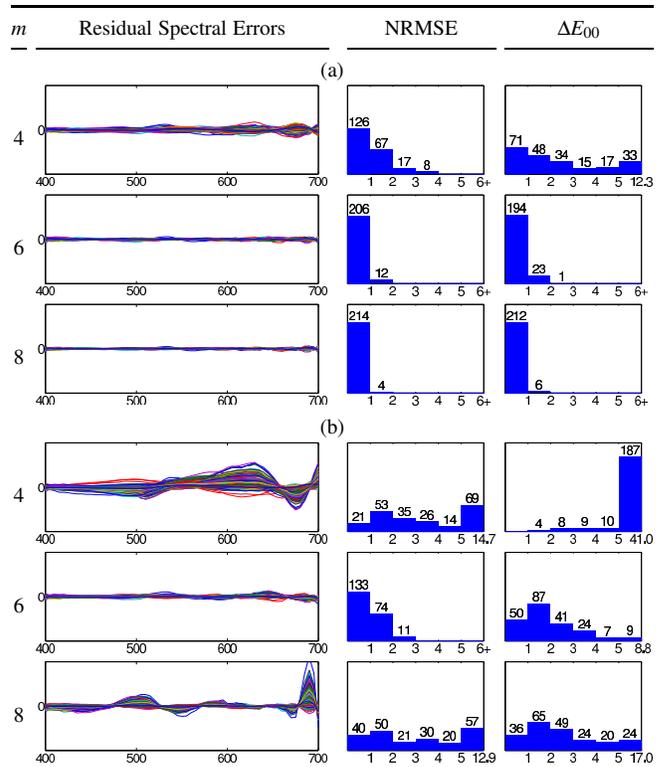


Figure 3. Errors and color differences incurred by computing Natural Colors/illuminant A color signals with a sharp basis found by (a) minimizing a weighted subspace projection and (b) optimizing the sharpness of \tilde{Q} its minimum angle.

The graphs in Fig. 4 show results obtained with the four other methods and the best method that optimizes sharpness. The first three rows show results obtained for color signal sets formed with a single illuminant; results shown in last four rows are obtained for sets formed with all three. In general, the lowest accuracy is attained using the method of Drew and Finlayson or by optimizing sharpness, while the highest accuracy is attained by minimizing the residual error explicitly or by minimizing a weighted subspace projection. Results obtained by PSO are not quite as accurate but are in every case as good as or better than those obtained by the two least accurate methods. Also notable is that the best sharpening intervals found by PSO are in every case far from evenly spaced or uniform. Furthermore, we learn from these results that accuracy can be very sensitive to interval choice; moving a single optimal interval boundary by the minimum of 5 nm increases average color difference in some cases by more than 600%.

Three trends are notable in Fig. 4. First, for D65 color signals, all methods, except that of Drew and Finlayson, yield similar accuracies. This result is likely due to the fact that D65 is close to white, for which the sharp basis approximation is exact and independent of \mathbf{T} . Second, the three most accurate methods have the desirable property of yielding accuracies that increase with dimensionality, a property not shared by the other methods. Fig. 3 shows a detailed example of this difference. We also see that in all cases the same accuracies are attained by minimizing the residual error and by minimizing a weighted subspace projection. In fact, these methods converge to the same \mathbf{T} matrices (up to column order) and thus to the same sharp bases, examples of which are shown in the top row of Fig. 2.

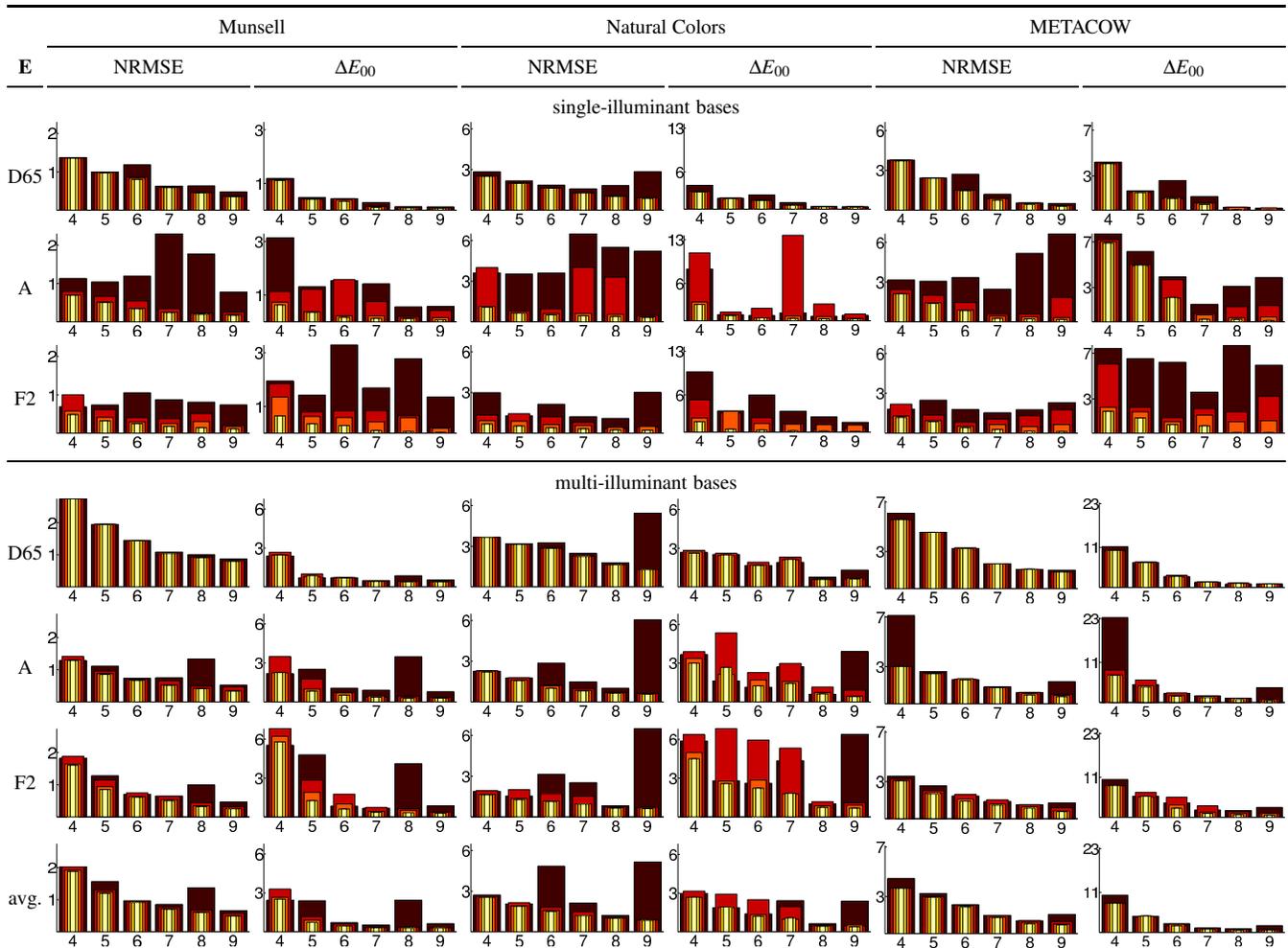


Figure 4. For dimensions $m = 4, 5, \dots, 9$, average NRMS errors and CIEDE2000 color differences of color signals computed with sharp bases constructed by: from outer (darker) to inner (lighter) bars, 1) evenly spaced sharpening intervals; 2) optimizing the sharpness of \hat{Q} and its minimum angle; 3) PSO of sharpening intervals; 4) minimizing a weighted subspace projection; and 5) explicitly minimizing the residual error.

Performance

Relative to the RGB cost, the theoretical cost of shading fragments with sharp basis coefficients is $\lceil \frac{m}{a} \rceil$, where a depends on the use of the alpha channel. In applications involving only opaque materials $a = 4$, because an RGBA vector can store four sharp basis coefficients; otherwise, $a = 3$. The cost of computing RGB coordinates is the cost of reading coefficients from the render texture(s) plus the cost of a $3 \times m$ matrix-vector multiplication at each pixel. Because these costs are likely to be dominated by the memory operation, we would expect that the total cost as a function of m would approximate a step function with a width of 3 or 4.

At any moment the performance of a real-time rendering application can be limited by computation on the CPU or vertex or fragment shading on the GPU. A conservative estimation of the real cost of rendering with sharp basis coefficients would therefore be the cost incurred by an application that: 1) requires minimal CPU computation; 2) renders as little geometry as possible; 3) has a simple lighting model; and 4) reads the reflectance properties of every fragment from GPU memory. Table 1 shows the costs, relative to RGB rendering, of three spectral rendering methods in an applica-

Table 1. Relative rendering costs.⁴

	61 point samples	m					
		4	5	6	7	8	9
RGB	1.00	27.6					
		orthonormal basis					
		3.21	6.51	23.0	70.4	124	259
		sharp basis					
		1.22	1.22	1.22	1.51	1.51	1.56

tion that meets these criteria, the rendering of a 1024×768 textured, full-screen quad with simple ambient lighting. Although the costs incurred by rendering with sharp basis coefficients shown here range from 22% to 56% above the cost of RGB rendering, the cost in a typical application should be significantly lower.

⁴Platform: Intel Core i7-970 CPU, Ubuntu 12.04 (AMD64, Linux kernel 3.5.0-40-generic), GeForce GTX 580 (NVIDIA driver version 295.53 with OpenGL 4.2.0). The alpha channel was reserved for transparency.

Discussion and Conclusion

The method of constructing \mathbf{T} by specifying evenly spaced sharpening intervals is unreliable. The approximations formed with the resulting bases are less accurate, and accuracy is not necessarily improved by simply increasing dimensionality. The same is true of methods that optimize sharpness. Furthermore, as shown in Fig. 1, sharpness of $\tilde{\mathbf{Q}}$ matters more than that of $\tilde{\mathbf{Q}}^+$, contrary to method of Drew and Finlayson, who sharpen the pseudoinverse. The accuracy of their method is likely a consequence of the implicit sharpening of $\tilde{\mathbf{Q}}$. Using PSO to find good sharpening intervals, which in general are neither uniform nor evenly spaced, can yield accurate results that do improve with dimensionality, but it is much less efficient than gradient-based NLP methods, as it involves m different $m \times m$ eigenvector problems and the evaluation of $p \cdot q$ residuals at every $2m$ -dimensional point on the trajectory of each particle in a swarm.

In this work we optimize for spectral accuracy without considering its effect on color difference. This choice is necessary in some applications, for instance when rendering interference effects, such as iridescence. Nevertheless, a number of methods exist for improving the color accuracy attained with finite linear models, such as weighted PCA [13]. For applications that use spectral information exclusively to predict color, it would be useful to investigate these methods in combination with those described here.

The use of the CIEDE2000 color difference formula to compare pixels in moving images is questionable. One measure of video image difference is SV-CIELAB, an extension of CIELAB that addresses the effect of spatio-temporal structure on color appearance [10]. To provide a more meaningful measure of color accuracy, a number of scenes containing materials with different reflectance characteristics could be constructed, such as indoor or outdoor scenes containing flowers, manufactured materials or human skin, all illuminated by a variety of sources. Videos generated from these scenes could then be compared using SV-CIELAB to measure differences between sharp basis approximations and ground truth.

Spectral rendering is essential in computer graphics applications requiring accurate color prediction, examples of which are found in optics, architectural rendering, industrial design, manufacturing and product marketing. In non-photorealistic visualization, spectral rendering can exploit metamerism to increase information density by causing colors to change meaningfully with illumination [2]. The efficiency gained by rendering with sharp basis coefficients permits these applications to operate in real time at a cost that is only a fraction above that of RGB rendering, even when spectra are represented at resolutions higher than 5 nm, since the method's $O(m)$ complexity is independent of spectral sampling resolution. We believe that spectral rendering also has interesting and useful applications as a creative tool in contexts such as computer games or computer-generated imagery used for special effects in films and television, not only for the many more degrees of freedom that it offers in the design of lighting and reflectance properties, but also for the possibility of using the spectra of familiar lights and materials in an intuitive way as starting points for creative variations.

Acknowledgements

The author is grateful to the following people for discussions and suggestions: Michael H. Brill, Charles Card, Bruce Kapron, Paul Lalonde, Wu-Sheng Lu, Jan Meseth, George Tzanetakis and Brian Wyvill. This research was partially funded by the GRAND NCE as part of the PARALLEL project.

References

- [1] ASTM International. Publication No. E308-12: Standard practice for computing the colors of objects by using the CIE system, 2012.
- [2] S. Bergner, T. Moller, M. Tory and M. Drew. A practical approach to spectral volume rendering. *IEEE Trans. Vis. and Comput. Graphics*, 11(2):207–216, 2005.
- [3] C. F. Borges. Trichromatic approximation method for surface illumination. *J. Opt. Soc. Am. A*, 8(8):1319–1323, 1991.
- [4] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310(1):1–26, 1980.
- [5] Central Bureau of the Commission Internationale de l'Éclairage (CIE). Publication No. 142-2001: Improvement to industrial colour-difference evaluation, 2001.
- [6] M. S. Drew and G. D. Finlayson. Spectral sharpening with positivity. *J. Opt. Soc. Am. A*, 17(8):1361–1370, 2000.
- [7] M. S. Drew and G. D. Finlayson. Multispectral processing without spectra. *J. Opt. Soc. Am. A*, 20(7):1181–1193, 2003.
- [8] M. D. Fairchild and G. M. Johnson. METACOW: A public-domain, high-resolution, fully-digital, noise-free, metameric, extended-dynamic-range, spectral test target for imaging system analysis and simulation. In *IS&T/SID 12th Color Imaging Conference*, pp. 239–245, 2004.
- [9] G. D. Finlayson, M. S. Drew and B. V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *J. Opt. Soc. Am. A*, 11(5):1553–1563, 1994.
- [10] K. Hirai, T. Mikami, N. Tsumura and T. Nakaguchi. Measurement and modeling of chromatic spatio-velocity contrast sensitivity function and its application to video quality evaluation. In *IS&T/SID 18th Color Imaging Conference*, pp. 86–91, 2010.
- [11] J. Kennedy and R. Eberhart. A new optimizer using particle swarm theory. In *Proc. Sixth Intl. Symposium on Micro Machine and Human Science*, pp. 39–43, 1995.
- [12] O. Kohonen, J. Parkkinen and T. Jääskeläinen. Databases for spectral color science. *Col. Res. and Appl.*, 31(5):381–390, 2006.
- [13] H. Laamanen, T. Jetsu, T. Jääskeläinen and J. Parkkinen. Weighted compression of spectral color information. *J. Opt. Soc. Am. A*, 25(6):1383–1388, 2008.
- [14] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am. A*, 61(1):1–11, 1971.
- [15] J. Lehtonen, J. Parkkinen and T. Jääskeläinen. Optimal sampling of color spectra. *J. Opt. Soc. Am. A*, 23(12):2983–2988, 2006.
- [16] M. Mahy, L. V. Eycken and A. Oosterlinck. Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. *Col. Res. and Appl.*, 19(2):105–1121, 1994.
- [17] D. H. Marimont and B. A. Wandell. Linear models of surface and illuminant spectra. *J. Opt. Soc. Am. A*, 9(11):1905–1913, 1992.
- [18] C. S. McCamy, H. Marcus and J. G. Davidson. A Color-Rendition Chart. *J. App. Photog. Eng.*, 2(3):95–99, 1976.
- [19] M. Stokes, M. D. Fairchild and R. S. Berns. Precision requirements for digital color reproduction. *ACM Trans. Graphics*, 11(4):406–422, 1992.
- [20] A. Wilkie and A. Weidlich. A robust illumination estimate for chromatic adaptation in rendered images. *Computer Graphics Forum*, 29(4):1101–1109, 2009.

Author Biography

The author is currently a Ph.D. candidate in the Department of Computer Science at the University of Victoria. In 2009 he received from the same university B.Sc. degrees with distinction in Combined Physics and Biochemistry and Combined Computer Science and Mathematics.