Optimal Color Filters in the Presence of Noise

M. J. Vrhel, *Member, IEEE,* and H. J. Trussell, *Fellow, IEEE*

Abstract—**In this paper, the effect of noise on the number of effective channels (color filters) used to record a color image is investigated. Transmittances of color filters are calculated that minimize the mean square error that occurs when estimating, from the recorded data, the colors in the image under a collection of viewing illuminants. Since the results indicate that a significant improvement in color correction accuracy is achieved by using four channels, there is good reason to consider using four-tuples for representation of colorimetric information.**

I. Introduction

In many applications, it is desirable to obtain colorimetric information about an object for multiple viewing illuminants from measurement with a single device. Applications where this would be useful include the textiles industry, in which color variation of dye lots for several illuminants should be closely monitored, and the printing industry in which accurate color rendition of a product in a retail catalog is important for several illuminants. In applications such as art archival, it is clearly imperative to obtain accurate color measurements of the original image.

In the reproduction of a color image, the original image is initially recorded using a device that measures the reflected or transmitted energy of the image in a number of different wavelength bands. In practice, the color values of the original image under a particular viewing illuminant are estimated from the data obtained from the recording device. This estimated data is then used by a reproduction device to match the original image under that viewing illuminant. Typically there will be a small number of viewing illuminants under which the original image and the reproduction may be compared.

One approach to improve the color accuracy of a color measuring device is to use an increased number of color filters (channels) [10], [13], [19]. A method for computing the optimal transmittances of such color filters is demonstrated in [19]. As the number of color filters is increased, additional information about the reflectance spectra in the original image is obtained. This additional information is used to improve the color match with the original image under a particular viewing illuminant.

Neglected in previous work is the effect of noise on the optimal number of channels and on the optimal color filter transmittances. The noise in this case arises in the measurement process [9]. From the noise free analysis, the conclusion is that improved accuracy is always obtained by increasing the number of channels. However, depending upon the signal-to-noise ratio (SNR), it is possible that improved performance may be achieved using a reduced number of channels.

Currently, there exists a wide range of color scanners. Most low-end devices such as color desk-top scanners use three channels. Many high-end scanners use more than three channels. The SNR of a common eight bit/channel desk-top scanner is in the range of 35–40 dB. While high-end scanners usually digitize to 12 bit/channel, the SNR is limited by the photomultiplier tube and is in the range of 50–60 dB.

There is limited discussion in the literature pertaining to the effect of noise on optimal color filter selection. In [6], simulated recorded data for a collection of noise free optimal filter sets was perturbed and transformed to a perceptual color space. The filter set that produced the smallest perceptual color error was selected. This trial and error approach provides limited insight into the filter selection problem.

In this paper, an analytical solution to the filter selection problem is derived. The relationship of SNR to the optimal number of channels and to the optimal filter transmittances is investigated. This relationship will be useful in the design of high-end and low-end color scanners. In Section II, mathematical notation, color representation, and noise sources in imaging devices are discussed. Following this background information, the problem of selecting the color filter transmittances is formulated. In addition, the solution to the problem, which is derived in Appendix A, is discussed. In Section III, the physical constraint of filter nonnegativity is discussed. Finally, in Section IV a simulation using optimal sets of filters at various SNR's is performed.

II. Problem Formulation and Solution

In this section, the CIE XYZ color space and the sources of noise in imaging systems are first reviewed. Following this background information, the problem of selecting an optimal set of color filters in the presence of noise is formulated and the solution is discussed.

A. Mathematical Notation and Color Background

A vector space notation for color systems will be used in the problem development [3]–[5], [7], [17], and [23]. The vector space approach allows straightforward formulation of problems in color reproduction. In the vector space approach, the visible spectrum (400–700 nm) is sampled at N wavelengths. Sampling requirements for color signals are discussed in [18].

If P color filters are used to obtain Pdifferent measurements, then in the vector space notation the recording process

Manuscript received June 21, 1993; revised April 13, 1994. The associate editor coordinating the review of this paper and approving it for publication was Dr. Fredrick Mintzer

M. J. Vrhel is with the Biomedical Engineering and Instrumentation Program, National Center for Research Resources, National Institutes of Health, Bethesda, MD 20892 USA.

H. J. Trussell is with the Electrical and Computer Engineering Department, North Carolina State University, Raleigh, NC 27695-7911 USA.

IEEE Log Number 9411141.

is represented using

$$\mathbf{c}_j = \mathbf{N}^T \mathbf{L}_r \mathbf{f}_j + \mathbf{u}_j \tag{1}$$

where $\mathbf{c}_j$ is the recorded P-stimulus vector at pixel j, $\mathbf{u}_j$ is additive noise, $\mathbf{f}_j$ is an N-dimensional vector representing the spectral reflectance (or transmittance) of an object at pixel j, $\mathbf{L}_r$ is an $N \times N$ diagonal matrix representing the spectral power distribution of the recording illumination, and $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_P]$ where $\mathbf{n}_i$ represents the spectral transmittance of the ith filter.

The vector notation can be used to express the CIE XYZ method of color quantification. If the matrix $\mathbf{A}$ contains the CIE 1931 XYZ color matching functions, then the CIE tristimulus vector

$$\mathbf{t}_{ij} = \mathbf{A}^T \mathbf{L}_i \mathbf{f}_j \tag{2}$$

quantifies the color of the object with spectral reflectance $\mathbf{f}_j$ under illuminant $\mathbf{L}_i$ [24]. It has been shown that the color of an object with spectral reflectance $\mathbf{f}_j$ under illuminant $\mathbf{L}_i$ is uniquely determined by the orthogonal projection of $\mathbf{f}_j$ onto the range space of $\mathbf{L}_i\mathbf{A}$ [4]. For this reason the range space of the matrix $\mathbf{L}_i\mathbf{A}$ is referred to as a human visual illuminant subspace (HVISS).

B. Noise Background

In electronic imaging devices, the photodetector is the element that converts optical radiation to electrical current. There are two main classes of photodetectors used for measuring color images. Solid state sensors include CCD and photo diode arrays that are inexpensive but noisy. Photomultiplier tubes are used in high-end flat-bed and drum scanners where more precise measurement is required. This increased accuracy is purchased at a substantial increase in price.

In any electronic device, there will be thermal noise present. Thermal noise is caused by random electron fluctuations in the device elements and can be well modeled as additive Gaussian signal-independent noise. Besides the effects of thermal noise on the measurement of optical radiation, the measurement will exhibit uncertainty due to the quantum nature of light. The measurement process can be considered a Poisson counting process since the sensor is in effect counting the arrival of photons. This type of noise is often referred to as shot noise. If the number of photons detected by the device is large, then the signal can be well modeled by a signal-dependent Gaussian distribution. Since the underlying process is Poisson, the variance of the Gaussian approximation is equal to the mean. Solid state detectors that are used at fairly high illuminance levels are most affected by this type of noise. At low illuminance levels, thermal noise is a major factor. If the gain of the photomultiplier tubes is high, it can be the source of additional shot noise.

A general noise model for a color imaging system can be expressed as follows:

$$\mathbf{c} = \mathbf{N}^T \mathbf{L}_r \mathbf{f} + g(\mathbf{c}) \circ \mathbf{u}_1 + \mathbf{u}_2 \tag{3}$$

where $\mathbf{u}_1 = [u_{11}, u_{12}, u_{13}]^T$, $g(\mathbf{c}) = [g(c_1), g(c_2), g(c_3)]^T$, and

$$g(\mathbf{c}) \circ \mathbf{u}_1 = [g(c_1)u_{11}, g(c_2)u_{12}, g(c_3)u_{13}]^T.$$

The term $\mathbf{u}_2$ represents the thermal noise and therefore has a zero mean Gaussian distribution. The term $g(\mathbf{c}) \circ \mathbf{u}_1$ represents the shot noise. The term $\mathbf{u}_1$ has a zero mean Gaussian distribution, and the function $g(\mathbf{c})$ models the signal dependence of the shot noise. In the case of Poisson noise, if the signal does not have a large dynamic range the noise can be well-approximated by a signal-independent model.

In this paper, the measurement noise is approximated as signal-independent. This is a reasonable assumption if the range of $g(\mathbf{c})$ is small relative to the variance of $\mathbf{u}_2$. Reference [1] contains a noise analysis of a vidicon camera. The vidicon camera, which uses an ionization sensor similar to a CCD, was shown to be dominated by signal-independent noise. There is some useful discussion of thermal and shot noise in [14]. A formulation of the problem of selecting filters in the presence of signal-dependent noise is contained in [20].

C. Optimal Color Filters

When designing a color imaging device there is usually limited control over the sensor characteristics. The overall system spectral sensitivity is obtained by careful selection of the color filters and recording illuminant. The problem of how to select the color filters and recording illuminant can be formulated as an optimization problem. Assume there are K viewing illuminants. The original image may be compared to the reproduction under any one viewing illuminant. From data obtained with a single set of color filters, the CIE XYZ tristimulus vectors of the original image under each viewing illuminant are estimated. The cost function is the mean square error (MSE) between the tristimulus vectors which describe the colors of the original image under each of the K viewing illuminants and the estimated tristimulus vectors. The MSE approach has produced excellent results in noise free simulations [19].

Mathematically the cost function, which is minimized with respect to the color filters/recording illuminant matrix $\mathbf{LN}$, is

$$\varepsilon = \sum_{i=1}^{K} \eta_i^2 E\{||\mathcal{P}_i(\mathbf{c}) - \mathbf{O}_i^T \mathbf{f}||^2\} \tag{4}$$

where $\mathcal{P}_i$ is the illumination color correction transformation for the ith illuminant, $\mathbf{c}$ is the data obtained from the output of the color filters, the η_i^2 are used to provide a weighting for the cost of color errors under the various illuminants, and the expectation operator is taken over the reflectance spectra $\mathbf{f}$ and the additive noise. The $N \times 3$ dimensional matrices $\mathbf{O}_i \quad i = 1 \cdots K$ are constructed to have orthonormal columns with the range space of $\mathbf{O}_i$ the same as the range space of $\mathbf{L}_i\mathbf{A}$. The orthonormalization is performed to remove weightings caused by magnitude differences in the illuminants. Note that $\mathbf{O}_i^T\mathbf{f}$ quantifies the color of spectral reflectance $\mathbf{f}$ under illuminant $\mathbf{L}_i$. For simplicity the color filters/recording illuminant matrix will be denoted by $\mathbf{G} = \mathbf{LN}$.

It can be shown that the linear minimum MSE estimator of the three dimensional vectors $\mathbf{d}_{ij} = \mathbf{O}_i^T \mathbf{f}_j$ is

$$\begin{aligned}\hat{\mathbf{d}}_{ij} &= \mathcal{P}_i(\mathbf{c}_j) \\ &= \mathbf{O}_i^T \mathbf{K}_f \mathbf{G}[\mathbf{G}^T \mathbf{K}_f \mathbf{G} + \mathbf{K}_u]^{-1} \\ &\quad \cdot [\mathbf{c} - \bar{\mathbf{u}} - \mathbf{G}^T \bar{\mathbf{f}}] + \mathbf{O}_i^T \bar{\mathbf{f}} \end{aligned} \tag{5}$$

where $\mathbf{K}_f$ is the covariance matrix of the reflectance spectra, $\mathbf{K}_u$ is the covariance matrix of the noise, $\bar{\mathbf{f}}$ is the mean of the reflectance spectra, and $\bar{\mathbf{u}}$ is the mean of the noise [14]. The value $\mathbf{t}_{ij}$ can be obtained from $\mathbf{d}_{ij}$ by a linear transformation. The sensor measurements for the P channels are performed independently on similar channels. Therefore, it is assumed that $\mathbf{K}_u = \mathbf{I}\sigma^2$.

The total power collected by the recording device is the integral of the image intensity over the duration of the measurement. Since the intensity of the image and the measurement time are limited by practical considerations, it is necessary to include a constraint on the signal power. The constraint on the signal power can be expressed as

$$E\{||\mathbf{G}^T \mathbf{f}||^2\} = \rho$$

where ρ is a constant. With a constraint on the signal power, the SNR in each channel will decrease as the number of channels is increased. This decrease in SNR occurs for devices that simultaneously record all P channels, and those that record one channel at a time.

Devices that simultaneously record all P channels include those with CCD color filter arrays, dichroic filters, or beam splitters. For these devices, an increase in the number of channels divides the signal energy among a larger number of channels while the level of noise in each channel remains the same. This results in a decreased SNR in each channel as the number of channels is increased.

Devices that record one channel at a time include those with color filter wheels. For these devices, an increase in the number of channels divides the measurement time over a larger number of channels. As the number of channels increases, the decrease in measurement time per channel will result in a decreased SNR for each channel.

From the results of Appendix A, an optimal set of filters that minimize ε will be contained in the set

$$C_G = \{\mathbf{G} \,|\, \mathbf{G} = \mathbf{K}_f^{-1/2} \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{V}^T \quad \mathbf{V}^T = \mathbf{V}^{-1}\} \tag{7}$$

where the columns of $\mathbf{U}^*$ are the P eigenvectors associated with the P largest eigenvalues of the matrix $\mathbf{K}_f^{1/2} \mathbf{S}\mathbf{S}^T \mathbf{K}_f^{1/2}$, $\mathbf{S} = [\eta_1 \mathbf{O}_1, \eta_2 \mathbf{O}_2, \cdots, \eta_K \mathbf{O}_K]$, $\Lambda^* = \text{Diag}\,[\lambda_1^*, \cdots \lambda_P^*]$, $\lambda_i^* = \sqrt{\gamma_i^*}$, and the γ_j^* are calculated using the results from Appendix A as follows:

1) From (A-26) calculate $\tilde{\gamma}_{r,r}$ for $r = 1, \cdots, P$.
2) The largest value of r for which $\tilde{\gamma}_{r,r} > 0$ is M, the number of nonzero γ_i^*, and

$$\gamma_j^* = \tilde{\gamma}_{j,M} \quad j = 1, \cdots, M \quad \gamma_j^* = 0 \quad j = M+1, \cdots, P.$$

It is of interest to consider the asymptotic behavior of the optimal filters with respect to the noise power. From (A-26), the γ_j in the noiseless case are selected as

$$\gamma_j^* = \kappa \frac{\sqrt{\delta_j}}{\sum_{i=1}^{P} \sqrt{\delta_i}} \quad j = 1, \cdots, P. \tag{8}$$

If the signal energy is equally distributed in the P dimensional space, in other words if $\delta_1 = \delta_2 = \cdots = \delta_P$, then the γ_j^* will not change as the SNR, (κ/σ^2), decreases.

If the signal energy is not equally distributed in the P-dimensional space, in other words if $\delta_1 > \delta_2 > \cdots > \delta_P$, then as the SNR decreases, the signal obtained in the subspace associated with δ_P will eventually be corrupted with noise to the point that the optimal value γ_P^*, obtained from (A-26), is zero. If γ_P^* is zero, then the Pth singular value, λ_P^*, of the matrix $\mathbf{H} = \mathbf{K}_f^{1/2} \mathbf{G}$ is zero which implies that the filter matrix $\mathbf{G} = \mathbf{L}_r \mathbf{N}$ is of rank $P - 1$ (see (A-1)-(A-5)). In this case, using more than $P - 1$ filters results in an increased MSE as opposed to using the optimal $P - 1$ filters. As the noise power becomes large, (κ/σ^2) will be less than $(\sum_{i=1}^{2} \sqrt{\delta_i}/\sqrt{\delta_2}) - 2$ (provided $\delta_1 > \delta_2$), and using more than one filter will result in an increased MSE.

III. Physical Constraints

Since a set of realizable color filters must have nonnegative transmittances, an optimal set of realizable filters will be any set in the intersection of C_G and $C_n = \{\mathbf{G} \,|\, \mathbf{G} \geq \mathbf{0}\}$. The parameter that could be adjusted to find a matrix $\mathbf{G}$ that satisfies both constraints is the unitary matrix $\mathbf{V}^T$. Explicitly the goal is to find a unitary matrix $\mathbf{V}$ such that matrix $\mathbf{K}_f^{-1/2} \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{V}^T$ is nonnegative. A unitary transformation $\mathbf{V}$ that results in a nonnegative set of filters may not exist, in which case the intersection of C_G and C_n is empty. If the intersection is empty, then it is of interest to find a matrix $\mathbf{G}$ in C_n that is a minimum distance from C_G with respect to some measure. Finding such a matrix is difficult due to the nonconvexity of C_G. An easier problem, which may result in a suboptimal solution is the following: find a matrix $\mathbf{V}$ that is a minimum distance from the set of unitary matrices subject to a nonnegativity constraint on the matrix $\mathbf{G} = \mathbf{K}_f^{-1/2} \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{V}^T$. This approach can be formulated as the following optimization problem.

Minimize

$$\frac{\sigma_{\max}(\mathbf{V})}{\sigma_{\min}(\mathbf{V})} \tag{9}$$

with respect to $\mathbf{V}$, subject to

$$\mathbf{K}_f^{-1/2} \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{V}^T \geq 0 \quad \text{Trace}\,[\mathbf{V} \mathbf{\Lambda}^* \mathbf{\Lambda}^* \mathbf{V}^T] = \kappa$$

where $\sigma_{\max}(\mathbf{V}), \sigma_{\min}(\mathbf{V})$ are the maximum and minimum singular values of the matrix $\mathbf{V}$, respectively. The constraint $\text{Trace}\,[\mathbf{V} \mathbf{\Lambda}^* \mathbf{\Lambda}^* \mathbf{V}^T] = \kappa$ forces the signal power to remain at a constant value. The minimum of the objective function is unity, and if that value is achieved, then the matrix $\mathbf{V}$

is unitary. Standard constrained optimization methods can be used to obtain a solution to the above problem [11].

Since the intersection of C_G and C_n could be empty, it may be more appropriate to impose the nonnegativity constraint on the optimization problem

Maximize

$$\text{Trace}\,[\mathbf{S}\mathbf{S}^T\mathbf{K}_f\mathbf{G}[\mathbf{G}^T\mathbf{K}_f\mathbf{G}+\mathbf{I}\sigma^2]^{-1}\mathbf{G}^T\mathbf{K}_f] \qquad (10)$$

with respect to G, subject to

$$\mathbf{G}^T \geq 0 \quad \text{Trace}\,[\mathbf{G}^T\mathbf{K}_f\mathbf{G}] = \kappa$$

which is equivalent to minimizing ε of (4) with respect to a nonnegative matrix G. The above optimization problem is difficult to solve and large scale nonlinear programming techniques may be necessary. The set of filters obtained from the optimization problem in (9) could be used as an initial starting point in a numerical solution to the above problem. An approach to enforce the nonnegativity constraint and reduce the number of parameters is to assume a parametric model for each filter. This approach was used by Vora and Trussell in the design of filters to span the HVISS [22].

IV. Experiment Results and Discussion

To test the performance of the color filters, experiments were performed in which the tristimulus vectors of reflectances under several illuminants were estimated from simulated recorded data obtained from a single set of filters at various SNR's. An ensemble of 343 spectral reflectances from a color copier were recorded using a spectraradiometer. A sampling width of 10 nm was used that resulted in $N = 31$ samples between 400 and 700 nm. Sets of three-, four-, five-, six-, and seven-color filters were calculated for the copier data set and the illuminants; CIE incandescent illuminant A, CIE daylight illuminant D65, and CIE fluorescent illuminant F2, at SNR's of 30, 35, 40, 45, 50, and ∞ dB. Since the accuracy of single precision, floating point arithmetic is only six–seven significant digits, the ∞ dB case actually corresponds to about 70 dB. Because of the similar results obtained in the 50 dB and ∞ dB cases, it was decided that a 60 dB series of experiments would not be necessary. Since three illuminants were used, $K = 3$ in (4). Uniform weighting of the errors under each viewing illuminant was used so that $\eta_i^2 = 1 \quad i = 1, \cdots, K$ in (4). Nonnegative sets of filters were calculated using the optimization problem in (9). The minimization was performed using a commercial scientific optimization subroutine.

Sets of filters greater than seven were not considered since most reflectance spectra can be accurately modeled using seven basis vectors [21]. The use of the copier data set provides results similar to those obtained from other spectral reflectance ensembles. In [19], optimal color filters for the noise free case are derived for a variety of spectral reflectance ensembles. In that work, similar results were achieved among the various data sets.

Of interest is how the optimal color filters varied with a change in the SNR. The most significant change occurred when using seven filters. At SNR's of 36.8 dB or lower the use of seven filters results in a larger MSE compared to using six filters. Therefore at SNR's or 36.8 dB or less the optimal seven filters are actually six filters. In most cases, the change in the shape of the nonnegative filters for a change in SNR was not dramatic. A typical change is demonstrated in Figs. 1–4. In Figs. 1 and 2, the six nonnegative filters at a 50 dB SNR are shown. The plots have values greater than one since they represent the filters combined with a recording illuminant. Figs. 3 and 4 contain the difference between the nonnegative six filters at a 30 dB SNR and those contained in Figs. 1 and 2. The fact that there is not a dramatic change is encouraging since it implies that the filter shapes need not be optimized for a specific SNR. Of importance, though, is the number of channels necessary to achieve a desired level of color accuracy. As was already noted, there exist situations in which the addition of filters results in decreased accuracy.

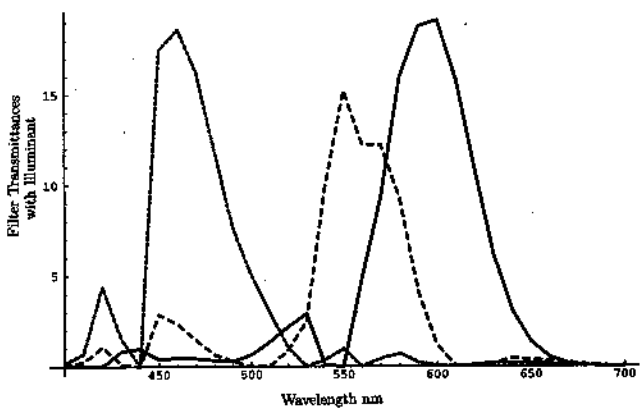

Fig. 1. Three of the nonnegative six filters designed for 50 dB SNR with the recording illuminant.

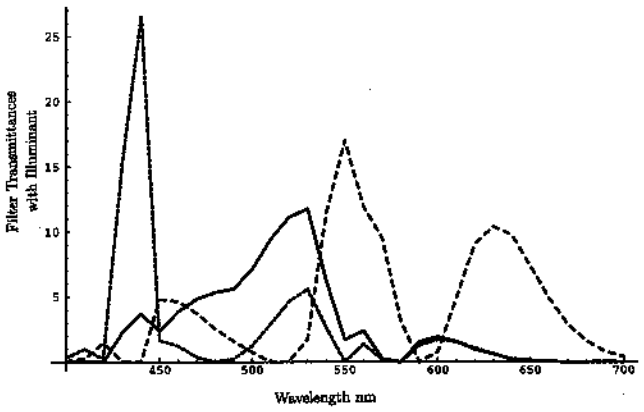

Fig. 2. The remaining three of the nonnegative six filters designed for 50 dB SNR.

Since the goal is accurate color recording, it is necessary to consider the color errors produced when recording with the filter sets at various SNR's. Of interest are the SNR's for which the use of additional filters provide little or no improvement in the color accuracy. To investigate the effect of noise on the color accuracy, simulated recording of the copier data set was performed using each of the filter sets at their designed SNR. The recorded data can be express as

$$\mathbf{c}_j = \mathbf{G}^T\mathbf{f}_j + \mathbf{u}_j \quad j = 1, 2 \cdots, 343 \qquad (11)$$

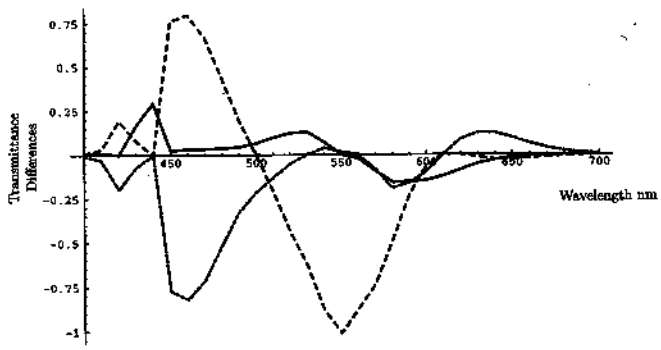

Fig. 3. Difference between the three filters in Fig. 1 and a set designed for 30 dB SNR.

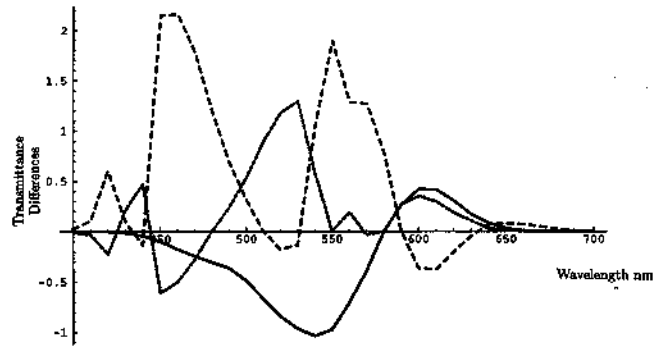

Fig. 4. Difference between the three filters in Fig. 2 and a set designed for 30 dB SNR.

where the matrix $\mathbf{G}^T$ denotes the filter set, $\mathbf{f}_j$ denotes spectral reflectance sample j, and $\mathbf{u}_j$ denotes the additive noise.

The goal is to estimate, from the recorded data, the CIE tristimulus vectors of the spectral reflectance samples under the three viewing illuminants. The CIE tristimulus vectors can be expressed as

$$\mathbf{t}_{ij} = \mathbf{A}^T\mathbf{L}_i\mathbf{f}_j \quad j = 1, 2, \cdots, 343 \quad i = 1, 2, 3. \tag{12}$$

The CIE tristimulus vectors were estimated from the recorded data using a LMMSE estimator. The CIE color difference measure ΔE^*_{ab} is often used as a measure of perceptual color difference [2], [24]. The average color tolerance accepted in printing applications has been studied and found to be approximately a ΔE^*_{ab} of six. The standard deviation in the accepted tolerance was 3.63 ΔE^*_{ab} [15]. In another study, the perceptibility tolerance for pictorial images was investigated and the average ΔE^*_{ab} was found to be 2.15 [16].

The ΔE^*_{ab} values between $\mathbf{t}_{ij}$ and the estimate $\hat{\mathbf{t}}_{ij}$ were calculated. Tables I-A–II-B contain the results for the simulations. The values in the tables are denoted as follows: ΔE_{avg}, the average ΔE^*_{ab} value of the set; $\Delta E_{\max}$, the maximum ΔE^*_{ab} in the set; $\Delta E^*_{ab} \geq 3$, the number of errors with a ΔE^*_{ab} greater than three; and MSE, the sum of the MSE's defined by (4). The number of ΔE^*_{ab} values greater than three are given since such values are perceptually noticeable by the average observer. For a particular viewing illuminant, the white point of the data set was the CIE tristimulus vector of the spectral reflectance of the paper from which the data set was measured. The nonnegative sets of filters are denoted as N3 to N7, and the optimal filters, which are not nonnegative, are denoted as

TABLE I-A
NOISE RESULTS OBTAINED WITH OPTIMAL FILTERS THAT ARE NOT CONSTRAINED TO BE NONNEGATIVE

SNR		30dB			35dB			40dB		
illum		D65	A	F2	D65	A	F2	D65	A	F2
O3	ΔE_{avg}	4.80	5.09	4.75	3.30	3.40	3.05	2.64	2.61	2.28
	ΔE_{max}	21.59	25.22	29.63	16.25	15.56	17.00	12.95	11.90	13.74
	$\Delta E^*_{ab} \geq 3$	226	217	215	150	143	134	109	108	75
	MSE	4.90E-3			2.25E-3			1.41E-3		
O4	ΔE_{avg}	4.32	4.65	4.48	2.44	2.66	2.68	1.40	1.57	1.73
	ΔE_{max}	16.76	17.54	19.02	10.32	8.95	10.29	6.35	5.04	6.30
	$\Delta E^*_{ab} \geq 3$	210	214	207	103	113	121	23	37	39
	MSE	4.14E-3			1.40E-3			5.36E-4		
O5	ΔE_{avg}	4.32	4.62	4.45	2.45	2.61	2.58	1.41	1.47	1.56
	ΔE_{max}	16.80	18.17	19.53	10.33	9.24	10.49	6.36	5.20	6.39
	$\Delta E^*_{ab} \geq 3$	201	209	204	103	106	111	23	33	31
	MSE	4.07E-3			1.31E-3			4.39E-4		
O6	ΔE_{avg}	4.32	4.63	4.41	2.44	2.60	2.48	1.37	1.47	1.39
	ΔE_{max}	16.43	18.16	18.42	9.74	9.30	8.89	5.56	5.11	5.27
	$\Delta E^*_{ab} \geq 3$	210	209	203	100	108	105	21	34	21
	MSE	4.05E-3			1.29E-3			4.08E-4		
O7	ΔE_{avg}	4.32	4.63	4.41	2.44	2.60	2.48	1.37	1.47	1.40
	ΔE_{max}	16.43	18.16	18.42	9.74	9.30	8.89	5.61	5.12	5.17
	$\Delta E^*_{ab} \geq 3$	210	209	203	100	108	105	21	34	21
	MSE	4.05E-3			1.29E-3			4.08E-4		

O3 to O7, where the number indicates the number of filters. The results obtained with the filters O3 to O7 are given since those results represent the best possible performance of any set of P filters at that SNR. While these filters can not be physically realized the results provide a useful limit.

The filters labeled $\mathbf{A}^T\mathbf{L}_i$ denote the use of filters that are ideal in the noiseless case. These filters represent the color matching functions combined with the viewing illuminant for which the data is being corrected. Recording with the filters $\mathbf{A}^T\mathbf{L}_i$ provides a direct measurement of the tristimulus values $\mathbf{t}_{ij}$.

When increasing from three to four filters, there is a large decrease in MSE and ΔE_{avg} at SNR's greater than 40 dB. This occurs for both the optimal filters and the nonnegative filters. At 30 dB, there is little improvement when a fourth filter is included. For the nonnegative filters, using more than three filters at 30 dB results in an increased MSE and ΔE_{avg} values. A plot of the optimal four filters for a 45 dB SNR are shown in Fig. 5.

To obtain the tristimulus vectors $\mathbf{t}_{ij}$, the filters $\mathbf{A}^T\mathbf{L}_i$ are often considered optimal. In the absence of noise, the filters $\mathbf{A}^T\mathbf{L}_i$ produce no color errors. In the presence of noise, however, the filters $\mathbf{A}^T\mathbf{L}_i$ are not optimal as evident when comparing N3 with $\mathbf{A}^T\mathbf{L}_i$ at 30 dB. In fact, the filters N3 perform as well, if not better than the filters $\mathbf{A}^T\mathbf{L}_i$ for SNR's less than 40 dB. This is a significant result and is often not considered by traditional color scientists.

Comparison of the nonnegative filters of Tables II-A and II-B with the optimal filters of Tables I-A and I-B indicates that the nonnegativity constraint results in significantly larger errors at 30 dB. One interesting value is the N5 filters at 30 dB in Table II-A. Compare this to filters N4 and N6 at 30 dB. The N5 filters perform poorly. This effect is not observed in Table I-A for the O4, O5, and O6 filters at 30 dB. The effect occurs for the nonnegative filters because they are not optimal. The optimization problem in expression (9) was used to find a set of filters that was nonnegative and close to the set of optimal filters C_G.

TABLE I-B
NOISE RESULTS OBTAINED WITH OPTIMAL FILTERS THAT ARE NOT CONSTRAINED TO BE NONNEGATIVE

SNR		45dB			50dB			Inf dB		
illum		D65	A	F2	D65	A	F2	D65	A	F2
O3	ΔE_{avg}	2.38	2.28	1.98	2.29	2.15	1.87	2.25	2.06	1.84
	ΔE_{max}	11.05	9.94	12.22	10.82	8.86	11.39	10.69	8.17	10.50
	$\Delta E^*_{ab} \geq 3$	96	94	57	95	81	54	91	80	55
	MSE	1.14E-3			1.06E-3			1.02E-3		
O4	ΔE_{avg}	0.84	0.99	1.26	0.55	0.69	1.06	0.35	0.49	0.95
	ΔE_{max}	4.00	3.18	4.28	2.64	2.41	3.83	1.34	2.20	3.93
	$\Delta E^*_{ab} \geq 3$	1	1	8	0	0	7	0	0	6
	MSE	2.62E-4			1.75E-4			1.35E-4		
O5	ΔE_{avg}	0.84	0.84	1.03	0.54	0.50	0.77	0.34	0.17	0.59
	ΔE_{max}	4.00	2.99	4.31	2.64	1.82	3.21	1.36	1.21	3.01
	$\Delta E^*_{ab} \geq 3$	1	0	3	0	0	1	0	0	1
	MSE	1.62E-4			7.40E-5			3.35E-5		
O6	ΔE_{avg}	0.77	0.83	0.78	0.44	0.47	0.44	3.1E-2	2.8E-2	2.12E-2
	ΔE_{max}	3.25	2.99	3.11	1.85	1.72	1.82	0.21	0.14	0.13
	$\Delta E^*_{ab} \geq 3$	1	0	1	0	0	0	0	0	0
	MSE	1.29E-3			4.11E-5			3.61E-5		
O7	ΔE_{avg}	0.77	0.82	0.78	0.44	0.46	0.44	8.62E-3	7.52E-3	4.55E-3
	ΔE_{max}	3.20	2.99	3.00	1.81	1.72	1.72	3.19E-2	3.02E-2	3.96E-2
	$\Delta E^*_{ab} \geq 3$	1	0	1	0	0	0	0	0	0
	MSE	1.29E-4			4.08E-5			1.40E-8		

TABLE II-A
NOISE RESULTS OBTAINED WITH FILTERS THAT WERE CONSTRAINED TO BE NONNEGATIVE, PLUS RESULTS FOR COLORIMETRIC OPTIMAL FILTERS

SNR		30dB			35dB			40dB		
illum		D65	A	F2	D65	A	F2	D65	A	F2
N3	ΔE_{avg}	7.29	7.39	7.37	4.41	4.41	4.28	3.12	3.03	2.87
	ΔE_{max}	31.65	34.43	43.42	21.96	19.13	23.85	16.38	13.84	16.56
	$\Delta E^*_{ab} \geq 3$	280	280	288	195	205	203	128	126	119
	MSE	9.37E-3			3.68E-3			1.86E-3		
N4	ΔE_{avg}	7.55	7.97	7.56	4.30	4.50	4.32	2.45	2.57	2.56
	ΔE_{max}	46.34	47.92	42.24	24.87	22.17	19.91	14.05	11.96	10.04
	$\Delta E^*_{ab} \geq 3$	288	284	280	195	203	200	96	103	104
	MSE	9.73E-3			3.25E-3			1.13E-3		
N5	ΔE_{avg}	8.78	8.71	8.55	5.23	5.03	5.05	2.99	2.84	2.88
	ΔE_{max}	45.14	45.62	70.99	26.24	25.36	30.05	15.01	13.94	15.25
	$\Delta E^*_{ab} \geq 3$	298	289	297	238	220	230	134	122	118
	MSE	1.15E-2			3.97E-3			1.29E-3		
N6	ΔE_{avg}	7.80	7.26	7.75	5.16	4.72	5.03	3.12	2.82	3.05
	ΔE_{max}	27.97	34.62	67.86	33.33	26.32	29.58	16.03	14.26	17.01
	$\Delta E^*_{ab} \geq 3$	263	274	274	222	214	212	130	118	132
	MSE	9.88E-3			3.91E-3			1.34E-3		
N7	ΔE_{avg}	7.80	7.26	7.75	5.16	4.72	5.03	3.12	2.82	3.05
	ΔE_{max}	27.97	34.62	67.86	33.33	26.32	29.58	16.03	14.26	17.01
	$\Delta E^*_{ab} \geq 3$	263	274	274	222	214	212	130	118	132
	MSE	9.88E-3			3.91E-3			1.63E-3		
$\mathbf{A}^T\mathbf{L}_i$	ΔE_{avg}	9.05	10.80	9.35	5.10	6.10	5.29	2.87	3.43	2.98
	ΔE_{max}	49.88	59.58	55.16	25.90	30.25	27.68	14.21	16.40	15.15
	$\Delta E^*_{ab} \geq 3$	276	308	287	212	257	221	117	161	126

In the tables, the SNR is defined as

$$10 \log \left[\frac{\kappa}{\sigma^2}\right]. \tag{13}$$

The SNR per channel is defined as

$$10 \log \left[\frac{\kappa}{P\sigma^2}\right] \tag{14}$$

where P channels are used in the system and κ is defined in (A-6). Therefore in the tables, a SNR of 45 dB is a SNR per channel of 40.23, 38.98, 38.01, 37.22, and 36.55 for three-, four-, five-, six-, and seven-channel systems, respectively. Quantization noise alone for an eight bit/channel system with a uniform quantizer results in a maximum SNR per channel of 48 dB. The values in the 45 dB column of the tables would be feasible for a standard desk-top scanner with eight bits/channel. Note that the four-channel system provides a significant improvement over the three-channel system at 45 dB. The trade-off for the colorimetric improvement is that the four-channel system is more complex and has a larger data storage requirement than the three-channel system.

V. CONCLUSION

The results of the simulation indicate that at SNR's greater than 40 dB, using four-color filters to obtain multiple viewing illuminant information provides a significant improvement over the three optimal filters. SNR's greater than 40 dB are feasible with the eight bit/channel system typically used in a desk-top scanner. For the SNR's found in high-end scanners, the addition of a fourth filter provides the greatest improvement. At 50 dB, using more than four filters provides little to no improvement. At SNR's less than 40 dB, the optimal three filters provide greater color accuracy for all three

TABLE II-B
NOISE RESULTS OBTAINED WITH FILTERS THAT WERE CONSTRAINED TO BE NONNEGATIVE, PLUS RESULTS FOR COLORIMETRIC OPTIMAL FILTERS

SNR		45dB			50dB			Inf dB		
illum		D65	A	F2	D65	A	F2	D65	A	F2
N3	ΔE_{avg}	2.61	2.40	2.30	2.29	2.15	1.87	2.25	2.06	1.84
	ΔE_{max}	12.59	10.55	14.41	10.82	8.86	11.39	10.69	8.17	10.50
	$\Delta E^*_{ab} \geq 3$	113	92	70	95	81	54	91	80	55
	MSE	1.28E-3			1.10E-3			1.02E-3		
N4	ΔE_{avg}	1.42	1.52	1.65	0.86	0.96	1.22	0.35	0.49	0.95
	ΔE_{max}	8.27	6.74	5.61	5.11	3.92	4.20	1.34	2.20	3.93
	$\Delta E^*_{ab} \geq 3$	23	31	40	2	2	12	0	0	6
	MSE	4.50E-4			2.35E-4			1.35E-4		
N5	ΔE_{avg}	1.61	1.57	1.69	0.93	0.83	1.09	0.34	0.17	0.59
	ΔE_{max}	7.42	6.87	7.33	3.42	3.72	4.60	1.36	1.21	3.01
	$\Delta E^*_{ab} \geq 3$	41	43	48	5	4	9	0	0	1
	MSE	4.45E-4			1.42E-4			3.35E-5		
N6	ΔE_{avg}	1.73	1.57	1.67	0.90	0.83	0.87	3.10E-2	2.80E-2	2.12E-2
	ΔE_{max}	10.61	9.30	11.14	4.87	3.75	3.95	0.21	0.14	0.13
	$\Delta E^*_{ab} \geq 3$	51	40	45	8	5	5	0	0	0
	MSE	4.01E-4			1.28E-4			3.61E-7		
N7	ΔE_{avg}	1.72	1.81	1.72	1.21	1.17	1.11	8.62E-3	7.52E-3	4.55E-3
	ΔE_{max}	8.03	8.13	9.63	5.04	5.33	5.39	3.19E-2	3.02E-2	3.96E-2
	$\Delta E^*_{ab} \geq 3$	53	64	52	20	22	15	0	0	0
	MSE	5.60E-4			2.15E-4			1.40E-8		
$\mathbf{A}^T\mathbf{L}_i$	ΔE_{avg}	1.61	1.93	1.67	0.91	1.08	0.94	0.0	0.0	0.0
	ΔE_{max}	7.91	9.34	8.86	4.48	5.29	4.92	0.0	0.0	0.0
	$\Delta E^*_{ab} \geq 3$	49	63	49	8	9	9	0	0	0

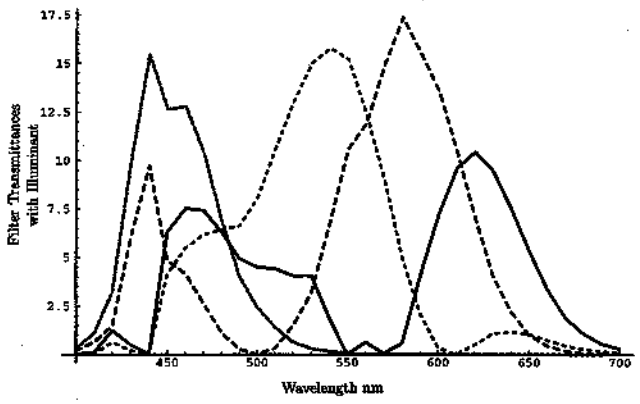

Fig. 5. The nonnegative four filters designed for 45 dB SNR.

illuminants, than do filters that provide direct measurement of the colorimetric data. At 50 dB using four filters provides results as good as those achieved by direct measurement with the filters $\mathbf{A}^T\mathbf{L}_i$. These results should be of interest to those involved in the design of color recording devices.

APPENDIX A
DERIVATION OF OPTIMAL COLOR FILTERS

Since the mean of the reflectance spectra is assumed known, and can be removed from the recorded data, it will be assumed for simplicity that the mean of the reflectance spectra ensemble is the zero vector. Substituting (5) into (4) and performing algebraic manipulations results in

$$\varepsilon = \text{Trace}\,[\mathbf{S}\mathbf{S}^T(\mathbf{K}_f - \mathbf{K}_f\mathbf{G} \cdot [\mathbf{G}^T\mathbf{K}_f\mathbf{G} + \mathbf{K}_u]^{-1}\mathbf{G}^T\mathbf{K}_f)]. \quad \text{(A-1)}$$

To produce a more manageable cost function, the matrix $\mathbf{H} = \mathbf{K}_f^{1/2}\mathbf{G}$ will be substituted into (A-1). Making the substitutions $\mathbf{H} = \mathbf{K}_f^{1/2}\mathbf{G}$ and $\mathbf{K}_u = \mathbf{I}\sigma^2$ into (A-1) produces

$$\varepsilon = \text{Trace}\,[\mathbf{S}\mathbf{S}^T(\mathbf{K}_f - \mathbf{K}_f^{1/2}\mathbf{H} \cdot [\mathbf{H}^T\mathbf{H} + \mathbf{I}\sigma^2]^{-1}\mathbf{H}^T\mathbf{K}_f^{1/2})]. \quad \text{(A-2)}$$

To find the optimal scanning filters, it is necessary to minimize ε with respect to the matrix $\mathbf{H}$. Equation (A-2) can be rewritten as

$$\varepsilon = \text{Trace}\,[\mathbf{S}\mathbf{S}^T\mathbf{K}_f] - \text{Trace}\,[\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}\mathbf{H} \cdot [\mathbf{H}^T\mathbf{H} + \mathbf{I}\sigma^2]^{-1}\mathbf{H}^T] \quad \text{(A-3)}$$

from which it is clear that an optimal matrix $\mathbf{H}$ that minimizes ε will maximize

$$\zeta = \text{Trace}\,[\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}\mathbf{H}[\mathbf{H}^T\mathbf{H} + \mathbf{I}\sigma^2]^{-1}\mathbf{H}^T]. \quad \text{(A-4)}$$

To algebraically simplify (A-4), the singular value decomposition of matrix $\mathbf{H}$ is substituted into (A-4) producing

$$\zeta = \text{Trace}\,[\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}\mathbf{U}\mathbf{\Lambda}[\mathbf{\Lambda}^2 + \mathbf{I}\sigma^2]^{-1}\mathbf{\Lambda}\mathbf{U}^T] \quad \text{(A-5)}$$

where $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T, \mathbf{U}^T\mathbf{U} = \mathbf{I}_{pxp}, \mathbf{\Lambda} = \text{Diag}\,[\lambda_1, \cdots, \lambda_2]$, and $\mathbf{V}^T = \mathbf{V}^{-1}$. Note that ζ is independent of the unitary matrix $\mathbf{V}$.

The constraint on the signal power results in a constraint on the $\lambda_i \quad i = 1, \cdots, P$. This follows from

$$\begin{aligned} E\{||\mathbf{G}^T\mathbf{f}||^2\} &= \text{Trace}\,[\mathbf{G}^T\mathbf{K}_f\mathbf{G}] + ||\mathbf{G}^T\bar{\mathbf{f}}||^2 \\ &= \text{Trace}\,[\mathbf{H}^T\mathbf{H}] + ||\mathbf{G}^T\bar{\mathbf{f}}||^2 \\ &= \text{Trace}\,[\mathbf{\Lambda}^2] + ||\mathbf{G}^T\bar{\mathbf{f}}||^2 = \rho \end{aligned}$$

which implies that

$$\text{Trace}\,[\mathbf{\Lambda}^2] = \rho - ||\mathbf{G}^T\bar{\mathbf{f}}||^2 = \kappa \quad \text{(A-6)}$$

where κ represents the power in the unknown portion of the signal.

The optimization problem can be written as follows.
Maximize

$$\zeta = \text{Trace}\,[\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}\mathbf{U}\mathbf{\Lambda}[\mathbf{\Lambda}^2 + \mathbf{I}\sigma^2]^{-1}\mathbf{\Lambda}\mathbf{U}^T] \quad \text{(A-7)}$$

with respect to $\mathbf{U}$ and $\mathbf{\Lambda}$ subject to

$$\text{Trace}\,[\mathbf{\Lambda}^2] = \kappa$$
$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_{pxp}.$$

It can be shown that the constraints define a compact subset of the $NP + P$ dimensional space defined by the coefficients λ_j and u_{ij} $\quad i = 1, \cdots, N \quad j = 1, \cdots, P$. This ensures the existence of a solution to the maximization problem [12, pp. 39–40].

Using the inequality [8, p. 183]

$$\text{Trace}\,[\mathbf{XY}] \leq \sum_{i=1}^{N} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) \tag{A-8}$$

where $\sigma_i(X)$ is the ith singular value of the matrix $\mathbf{X}$ and the singular values are indexed in nonincreasing magnitude, it will be shown that an optimal value for the matrix $\mathbf{U}$ is the matrix containing the P eigenvectors associated with the P largest eigenvalues of the matrix $\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}$. Equation (A-7) can be written as

$$\zeta = \text{Trace}\,[\mathbf{U}^T\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}\mathbf{U}\mathbf{\Lambda}[\mathbf{\Lambda}^2 + \mathbf{I}\sigma^2]^{-1}\mathbf{\Lambda}]. \tag{A-9}$$

Let

$$\mathbf{X} = \mathbf{U}^T\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}\mathbf{U} \tag{A-10}$$

and

$$\mathbf{Y} = \mathbf{\Lambda}[\mathbf{\Lambda}^2 + \mathbf{I}\sigma^2]^{-1}\mathbf{\Lambda}. \tag{A-11}$$

Note that the matrix $\mathbf{Y}$ is diagonal and if the diagonal elements of $\mathbf{\Lambda}$ are in nonincreasing order, then the diagonal elements of $\mathbf{Y}$ are also in nonincreasing order. In this case,

$$\sigma_i(\mathbf{Y}) = y_{ii} \tag{A-12}$$

where y_{ii} denotes the ith diagonal element of $\mathbf{Y}$. Using relation (A-8) and the fact that $\mathbf{Y}$ is diagonal produces

$$\text{Trace}\,[\mathbf{XY}] \leq \sum_{i=1}^{N} \sigma_i(\mathbf{X})y_{ii} \tag{A-13}$$

where $y_{11} \geq y_{22} \geq \cdots \geq y_{NN}$. The equality in (A-13) and hence the maximum of ζ is achieved if the columns of the orthonormal matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N]$ are the eigenvectors of the matrix $\mathbf{X}$ and the eigenvectors are ordered such that the corresponding eigenvalues are in a nonincreasing order. In this case

$$\mathbf{X}\mathbf{u}_i = \mathbf{u}_i\sigma_i(\mathbf{X}) \tag{A-14}$$

and $\sigma_1(\mathbf{X}) \geq \cdots \geq \sigma_N(\mathbf{X})$.

With the optimal matrix $\mathbf{U}$ known, the optimization problem of (A-7) can be written as

$$\text{Maximize} \sum_{i=1}^{P} \delta_i \frac{\lambda_i^2}{\lambda_i^2 + \sigma^2} \tag{A-15}$$

with respect to the λ_i, subject to

$$\sum_{i=1}^{P} \lambda_i^2 = \kappa$$

where the δ_i represent the eigenvalues of $\mathbf{K}_f^{1/2}\mathbf{S}\mathbf{S}^T\mathbf{K}_f^{1/2}$, and $\delta_1 \geq \cdots \geq \delta_N$.

An equivalent optimization problem that is easier to solve is found by substituting $\gamma_i = \lambda_i^2$ and constraining γ_i to be nonnegative. The optimization problem is

$$\text{Maximize} \sum_{i=1}^{P} \delta_i \frac{\gamma_i}{\gamma_i + \sigma^2} \tag{A-16}$$

with respect to the γ_i subject to

$$\sum_{i=1}^{P} \gamma_i = \kappa \quad \gamma_i \geq 0.$$

Making the following definitions

$$\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_P]^T$$
$$f[\boldsymbol{\gamma}] = \sum_{i=1}^{P} \delta_i \frac{\gamma_i}{\gamma_i + \sigma^2}$$
$$h[\boldsymbol{\gamma}] = \sum_{i=i}^{P} \gamma_i - \kappa$$
$$\mathbf{g}[\boldsymbol{\gamma}] = \boldsymbol{\gamma}$$
$$\mathbf{g}[\boldsymbol{\gamma}] = [g_1[\boldsymbol{\gamma}], \cdots, g_P[\boldsymbol{\gamma}]]^T$$

the above optimization problem becomes

$$\text{Maximize}\, f[\boldsymbol{\gamma}] \tag{A-17}$$

with respect to $\boldsymbol{\gamma}$ subject to $h[\boldsymbol{\gamma}] = 0$ and $\mathbf{g}[\boldsymbol{\gamma}] \geq \mathbf{0}$.

The Kuhn–Tucker optimality conditions are [11]

$$\nabla_{\boldsymbol{\gamma}} f[\boldsymbol{\gamma}^*] + \sum_{i=1}^{P} \nu_i \nabla_{\boldsymbol{\gamma}} g_i[\boldsymbol{\gamma}^*] + \chi \nabla_{\boldsymbol{\gamma}} h[\boldsymbol{\gamma}^*] = \mathbf{0}$$
$$\mathbf{g}[\boldsymbol{\gamma}^*] \geq \mathbf{0}$$
$$h[\boldsymbol{\gamma}^*] = 0$$
$$\boldsymbol{\nu}^T\mathbf{g}[\boldsymbol{\gamma}^*] = 0$$
$$\boldsymbol{\nu} \geq \mathbf{0}.$$

where χ and ν are Lagrange multipliers. Substituting in the definitions of f, h, and $\mathbf{g}$ into the above equations produces the first-order necessary conditions

$$\frac{\delta_j\sigma^2}{(\gamma_j + \sigma^2)^2} + \nu_j + \chi = 0 \quad \sum_{i=1}^{P} \gamma_i = \kappa$$
$$\nu_j \geq 0 \quad \gamma_j \geq 0 \quad \gamma_j\nu_j = 0 \quad j = 1, \cdots, P.$$

From the constraints, it is clear that at least one γ_j must be nonzero. In addition, if $\gamma_j > 0$, then $\nu_j = 0$, and if $\nu_j > 0$, then $\gamma_j = 0$. It remains to solve the above equations for the $\gamma_j = \lambda_j^2$.

For the γ_j, which are nonzero

$$\gamma_j = \sqrt{\frac{-\delta_j\sigma^2}{\chi}} - \sigma^2 \quad j = 1, \cdots, P \quad \gamma_j \neq 0. \tag{A-18}$$

Assuming there are M nonzero γ_j, summing over j for $\gamma_j \neq 0$ produces

$$\kappa = \sum_{\substack{j=1 \\ \gamma_j \neq 0}}^{P} \gamma_j = \frac{1}{\sqrt{-\chi}} \sum_{\substack{j=1 \\ \gamma_j \neq 0}}^{P} \sqrt{\delta_j \sigma^2} - M\sigma^2. \quad \text{(A-19)}$$

Solving for $(1/\sqrt{-\chi})$ produces

$$\frac{1}{\sqrt{-\chi}} = \frac{\kappa + M\sigma^2}{\sum_{i=1, \gamma_i \neq 0}^{P} \sqrt{\delta_i \sigma^2}}. \quad \text{(A-20)}$$

Substituting (A-20) into (A-18) results in

$$\gamma_j = (\kappa + M\sigma^2) \frac{\sqrt{\delta_j}}{\sum_{i=1, \gamma_i \neq 0}^{P} \sqrt{\delta_i}} - \sigma^2$$
$$j = 1, \cdots, P \quad \gamma_j \neq 0. \quad \text{(A-21)}$$

Equation (A-21) provides the values of the nonzero γ_j. It remains to show which γ_j are zero and which can be computed using (A-21).

Consider the $P - M$ values for which $\gamma_j = 0$. In this case, from the necessary conditions it follows that

$$\frac{\delta_j}{\sigma^2} + \nu_j + \chi = 0 \quad j = 1, \cdots, P \quad \gamma_j = 0. \quad \text{(A-22)}$$

Solving (A-20) for χ, substituting in the above, and solving for ν_j produces

$$\nu_j = \left(\frac{\sum_{i=1, \gamma_i \neq 0}^{P} \sqrt{\sigma^2 \delta_i}}{k + M\sigma^2} \right)^2 - \frac{\delta_j}{\sigma^2}. \quad \text{(A-23)}$$

If $\gamma_j = 0$ and hence $\nu_j \geq 0$, then from the above equation

$$k + M\sigma^2 \leq \sigma^2 \frac{\sum_{i=1, \gamma_i \neq 0}^{P} \sqrt{\delta_i}}{\sqrt{\delta_j}}. \quad \text{(A-24)}$$

From the above inequality, and (A-21) it follows that a term $\tilde{\gamma}_{j,M}$ can be defined as

$$\tilde{\gamma}_{j,M} = (\kappa + M\sigma^2) \frac{\sqrt{\delta_j}}{\sum_{i=1, \gamma_i \neq 0}^{P} \sqrt{\delta_i}} - \sigma^2 \quad j = 1, \cdots, P \quad \text{(A-25)}$$

where $\tilde{\gamma}_{j,M} = \gamma_j$ if $\tilde{\gamma}_{j,M} > 0$ and $\gamma_j = 0$ if $\tilde{\gamma}_{j,M} \leq 0$.

From (A-25) and the fact that

$$\frac{\sqrt{\delta_1}}{\sum_{\gamma_i \neq 0} \sqrt{\delta_1}} \geq \frac{\sqrt{\delta_2}}{\sum_{\gamma_i \neq 0} \sqrt{\delta_2}} \geq \cdots \geq \frac{\sqrt{\delta_P}}{\sum_{\gamma_i \neq 0} \sqrt{\delta_P}}$$

it follows that $\tilde{\gamma}_{1,M} \geq \tilde{\gamma}_{2,M} \geq \cdots \tilde{\gamma}_{P,M}$ and therefore that the M γ_j that are greater than zero must be those associated with the M largest δ_j.

Therefore, (A-25) can be rewritten as

$$\tilde{\gamma}_{j,M} = (\kappa + M\sigma^2) \frac{\sqrt{\delta_j}}{\sum_{i=1}^{M} \sqrt{\delta_i}} - \sigma^2 \quad j = 1, \cdots, P \quad \text{(A-26)}$$

and the optimal γ_j are

$$\gamma_j^* = \tilde{\gamma}_{j,M} \quad j = 1, \cdots, M \quad \gamma_j^* = 0 \quad j = M+1, \cdots, P.$$

A problem with (A-26) is that it depends on knowledge of M, the number of nonzero γ_j^*. Since this information is not known *a priori*, the question arises what happens if a value other than the number of positive γ_i^* is used for M in (A-26). It can be shown that if a value R that is different than the number of positive γ_i^* is used for M in (A-26), then the conditions $\tilde{\gamma}_{j,R} > 0 \quad j = 1, \cdots, R$ and $\tilde{\gamma}_{j,R} \leq 0 \quad j = R+1, \cdots, P$ are not satisfied [20].

From the uniqueness of M satisfying the above conditions, the inequalities $\tilde{\gamma}_{j,r} \leq \tilde{\gamma}_{j-1,r}, \tilde{\gamma}_{r+1,r+1} \leq \tilde{\gamma}_{r,r}$, and the fact that $\gamma_{r,r} \leq 0$ implies $\gamma_{r,r-1} \leq 0$ it follows that M is the largest value r for which $\tilde{\gamma}_{r,r} > 0$. Therefore, the γ_j^* values can be calculated by the following process:

1) From (A-26) calculate $\tilde{\gamma}_{r,r}$ for $r = 1, \cdots, P$.
2) The largest value of r for which $\tilde{\gamma}_{r,r} > 0$ is M, the number of nonzero γ_i^*, and

$$\gamma_j^* = \tilde{\gamma}_{j,M} \quad j = 1, \cdots, M \quad \gamma_j^* = 0 \quad j = M+1, \cdots, P.$$

An optimal set of filters that minimize ε given in (4) will be contained in the set

$$C_{Gn} = \{\mathbf{G} \,|\, \mathbf{G} = \mathbf{K}_f^{-1/2} \mathbf{U}^* \mathbf{A}^* \mathbf{V}^T \quad \mathbf{V}^T = \mathbf{V}^{-1}\} \quad \text{(A-27)}$$

where the columns of $\mathbf{U}^*$ are the P eigenvectors associated with the P largest eigenvalues of the matrix $\mathbf{K}_f^{1/2} \mathbf{S}\mathbf{S}^T \mathbf{K}_f^{1/2}$, $\mathbf{A}^* = \text{Diag}\,[\lambda_1^*, \cdots, \lambda_P^*]$, and $\lambda_i^* = \sqrt{\gamma_i^*}$.

REFERENCES

[1] R. A. Boie and I. J. Cox, "An analysis of camera noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 6, pp. 671–674, June 1992.

[2] CIE, "Recommendations on uniform color spaces, color-difference equations, psychometric color terms," *Supplement no. 2 of CIE Publ. no. 15.* (E-1.3.1) 1971, Bureau Central de la CIE, Paris, 1978.

[3] J. B. Cohen, "Color and color mixture: Scalar and vector fundamentals," *Color Res. Applicat.*, vol. 13, no. 1, pp. 5–39, Feb. 1988.

[4] J. B. Cohen and W. E. Kappauf,"Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks," *Amer. J. Psych.*, vol. 95, no. 4, pp. 537–564, Winter 1982.

[5] ——,"Color mixture and fundamental metamers: Theory, algebra, geometry, application," *Amer J. Psych.*, vol. 98, no. 2, pp. 171–259, Summer 1985.

[6] K. Engelhardt and P. Seitz, "Optimum color filters for CCD digital cameras," *Appl. Opt.*, vol. 32, no. 16, pp. 3015–3023, June 1993.

[7] B. K. P. Horn, "Exact reproduction of colored images," *Comput. Vision, Graph, Image Processing*, vol. 26, pp. 135–167, 1984.

[8] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis.* Cambridge, MA: Cambridge Univ. Pr., 1991.

[9] I. N. Jung, "Noise margins of color image drum scanners," *J. Photo. Sci.*, vol. 41, pp. 98–99, 1993.

[10] R. V. Kollarits and D. C. Gibbon, "Improving the color fidelity of cameras for advanced television systems," *SPIE Proc.*, vol. 1656, 1992.

[11] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison–Wesley, 1984.

[12] ——, *Optimization by Vector Space Methods.* New York, NY: Wiley, 1969.

[13] K. Martinez, J. Cupitt, and D. Saunders, "High resolution colorimetric imaging of paintings," *SPIE Proc.*, vol. 1901, 1993.
[14] W. K. Pratt, *Digital Image Processing*, 2nd ed. New York, NY: Wiley, 1991.
[15] S. Stamm, "An investigation of color tolerance," *TAGA Proceedings 1981*, pp. 156–173.
[16] M. Stokes, M. D. Fairchild, and R. S. Berns, "Precision requirements for digital color reproduction," *ACM Trans. Graphics*, vol. 11, no. 4, pp. 406–422, Oct. 1992.
[17] H. J. Trussell, "Applications of set theoretic methods to color systems," *Color Res. Appl.*, vol. 16, no. 1, pp. 31–41, Feb. 1991.
[18] H. J. Trussell and M. Kulkarni, "Sampling and processing of color signals," submitted to *IEEE Trans. Image Processing*
[19] M. J. Vrhel and H. J. Trussell, "Filter considerations in color correction," *IEEE Trans. on Image Processing*, vol. 3, no. 2, pp. 147–161, Mar. 1994.
[20] M. J. Vrhel, "Mathematical methods of color correction," Ph.D. dissertation, North Carolina State University, May 1993.
[21] M. J. Vrhel, R. Gershon, and L. Iwan, "The measurement and analysis of object reflectance spectra," *Color Research and Application*, vol. 19, no. 1, pp. 4–9, Feb. 1994.
[22] P. L. Vora and H. J. Trussell, "A mathematical method for designing a set of color scanning filters," *SPIE Proc.*, vol. 1912, 1993.
[23] B. A. Wandell, "The synthesis and analysis of color images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 1, pp. 2–13, Jan. 1987.
[24] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York, NY: Wiley, 1982.

Michael J. Vrhel (S'87–M'87) was born in St. Joseph, Michigan in 1964. He received the B.S. degree in electrical engineering from Michigan Technological University, Houghton, MI, in 1987. He received the M.S. and Ph.D. degrees in electrical engineering from North Carolina State University, Raleigh, NC, in 1989 and 1993, respectively.

He was the recipient of a Kodak Fellowship from Eastman Kodak Company, Rochester, NY from 1989 to 1993. During 1992 he researched problems in color reproduction at the Eastman Kodak Company. Currently, Dr. Vrhel is a National Research Council Associate at the National Institutes of Health, Biomedical Engineering and Instrumentation Program. His current research interests include color reproduction, signal restoration/reconstruction, and signal processing with wavelets.

Dr. Vrhel is an associate member of Sigma Xi and a member of Pi Mu Epsilon.

H. J. Trussell (S'75–M'76–SM'91–F'94) was born in Atlanta, GA, Feb. 3, 1945. He received the B.S. degree from Georgia Tech in 1967, the M.S. degree from Florida State in 1968, and the Ph.D. degree from the University of New Mexico in 1976. He joined the Los Alamos Scientific Laboratory, Los Alamos, NM, in 1969, where he began working in the image and signal processing in 1971. During 1978–1979, he was a visiting professor at Heriot-Watt University, Edinburgh, Scotland. In 1980, he joined the Electrical and Computer Engineering Department at North Carolina State University, Raleigh, NC. During 1988–1989, he was a visiting scientist at the Eastman Kodak Company in Rochester, NY. He has won the IEEE-ASSP Society Senior Paper Award (1986 with M. R. Civanlar) and the IEEE-SP Society Paper award (1993 with P. L. Combettes).

He is a past associate editor for the TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING and currently an associate editor for the SIGNAL PROCESSING LETTERS. He is a member and past chairman of the Image and Multidimensional Digital Signal Processing Committee of the Signal Processing Society of the IEEE. He edits the electronic newsletter published by this committee.

He is a member of Sigma Xi, Tau Beta Pi, Phi Kappa Phi, the Signal Processing Society, the Optical Society of America and the Intersociety Color Council. His current interests include signal restoration/reconstruction, color measurement and reproduction, and mathematical methods.