A Comparison of Computational Color Constancy Algorithms—Part I: Methodology and Experiments With Synthesized Data

Kobus Barnard, Vlad Cardei, and Brian Funt

***Abstract*—We introduce a context for testing computational color constancy, specify our approach to the implementation of a number of the leading algorithms, and report the results of three experiments using synthesized data. Experiments using synthesized data are important because the ground truth is known, possible confounds due to camera characterization and pre-processing are absent, and various factors affecting color constancy can be efficiently investigated because they can be manipulated individually and precisely.**

The algorithms chosen for close study include two gray world methods, a limiting case of a version of the Retinex method, a number of variants of Forsyth's gamut-mapping method, Cardei *et al.*'s neural net method, and Finlayson *et al.*'s Color by Correlation method. We investigate the ability of these algorithms to make estimates of three different color constancy quantities: the chromaticity of the scene illuminant, the overall magnitude of that illuminant, and a corrected, illumination invariant, image. We consider algorithm performance as a function of the number of surfaces in scenes generated from reflectance spectra, the relative effect on the algorithms of added specularities, and the effect of subsequent clipping of the data. All data is available on-line at http://www.cs.sfu.ca/~color/data, and implementations for most of the algorithms are also available (http://www.cs.sfu.ca/~color/code).

***Index Terms*—Algorithm, color by correlation, color constancy, comparison, computational, gamut constraint, neural network.**

I. Introduction

THE IMAGE recorded by a camera depends on three factors: the physical content of the scene, the illumination incident on the scene, and the characteristics of the camera. The goal of computational color constancy is to account for the effect of the illuminant, either by directly mapping the image to a standardized illuminant invariant representation, or by determining a description of the illuminant which can be used for subsequent color correction of the image. This has important applications such as object recognition and scene understanding, as well as image reproduction and digital photography (see [1] for additional overview).

Manuscript received December 17, 2000; revised May 1, 2002. This work was supported by the National Research Council of Canada (NSERC) and Hewlett-Packard Laboratories. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark S. Drew.

K. Barnard was with Simon Fraser University, Burnaby, BC, Canada. He is now with the Computer Science Science Division, University of California, Berkeley, CA 94720-1776 (e-mail: kobus@cs.berkeley.edu).

V. Cardei and B. Funt are with Simon Fraser University, Burnaby, BC, Canada (e-mail: cardei@cs.sfu.ca; funt@cs.sfu.ca).

Publisher Item Identifier 10.1109/TIP.2002.802531.

In this paper we discuss a number of the leading algorithms and characterize their performance using synthesized data (all data is available on-line [2], as are implementations for most of the algorithms [3]). Experiments using synthesized data are important because the ground truth is known, possible confounds due to camera characterization and pre-processing are absent, and factors affecting color constancy can be efficiently investigated because they can be manipulated individually and precisely. These factors include input characteristics such as specularities, camera behaviors such as pixel clipping, and the statistics of illuminant and surface reflectance occurrence in training data and in testing data. Understanding these factors under controlled conditions is a necessary first step toward dealing with them in images taken with a real camera. Furthermore, the methodology used in this work has been informed by our work with a large data set of real images [1], [4]. This has enabled us to make our experiments with synthetic data closely relevant to the problems faced with when real image data is encountered.

II. Approaches to Computational Color Constancy

The goal of computational color constancy is to find a nontrivial illuminant invariant description of a scene from an image taken under unknown lighting conditions. This is often broken into two steps. The first step is to estimate illuminant parameters, and then a second step uses those parameters to compute illumination independent surface descriptors [5]–[7]. These descriptors can be quite abstract, but here we simplify matters by specifying that the illumination invariant description is an image of the scene as if it were taken under a known, standard, canonical, light [8]. The choice of the canonical illuminant is somewhat arbitrary. For image reproduction applications it makes most sense to use an illuminant for which the camera is balanced, and this is the choice we have used.

We assume a diagonal model of illumination change which maps the image taken under one illuminant, to the image taken under another illuminant (e.g., the canonical), by scaling each channel independently. For concreteness, consider a scene with a white patch. Suppose that the camera response to the white patch under the unknown illuminant is (R^U, G^U, B^U), and that the response under the known, canonical, illuminant is (R^C, G^C, B^C). Then the response to the white patch can be mapped from the unknown case to the canonical case simply by scaling the three channels by R^C/R^U, G^C/G^U, and B^C/B^U

respectively. To the extent that this same scaling works for the other, nonwhite patches, we say that the diagonal model holds. The efficacy of the diagonal model is largely a function of the vision system sensors, specifically whether or not they are narrow band, and whether or not they overlap[1] [9]–[12]. In the case of the camera used for the present work, the diagonal model is a good approximation. If the diagonal model leads to large errors, then performance may be improved by using sensor sharpening [13], [14].

The variants of Forsyth's gamut mapping method [8] directly estimate the diagonal mapping from the input image to the "corrected" image. The other algorithms considered in this paper estimate the color of the illuminant as defined by the camera response to a pure white, or a projection thereof (chromaticity). Using the diagonal model we can easily convert between these two approaches. An estimate of the illuminant (R, G, B) can be used to correct an image. Conversely, an estimate of the diagonal map also gives us an estimate of the illuminant (R, G, B), found by applying the inverse transform to the (R, G, B) of white under the canonical illuminant.

Often we are most interested in the chromaticity of the illuminant and/or a correction for chromaticity only. We remind the reader that chromaticity is color normalized by overall magnitude-one chromaticity space is $(r, g) = (R/(R+G+B), G/(R+G+B))$. Chromaticity is often sufficient because an illuminant magnitude is often implicitly present. For example, when a picture is taken, either a human operator or some mechanism has often set the aperture to a reasonable value. Thus, a correction for chromaticity, which leaves the overall brightness the same, is often sufficient for image reproduction applications. A number of color constancy algorithms have been developed which work entirely in some chromaticity space [15]–[18], and much progress has been made by taking advantage of the simplifications afforded by this strategy. However, since these algorithms ignore the magnitude of the image pixels, they are potentially less powerful than algorithms which attempt to use information that may be implicit in those values. For example, it is commonly recognized that specular highlights carry information about the illuminant chromaticity [19]–[23], and the fact that they are relatively bright is of use to some algorithms. This means that it can be beneficial to use an algorithm which estimates three parameters even if the goal is chromaticity correction.

III. Error Measures

We consider six error measures. The first treats the illuminant (R, G, B), and the corresponding estimate thereof, as vectors in (R, G, B) space, and computes the angle between these two vectors in degrees. Specifically, if $\mathbf{T} = (R_T, G_T, B_T)$ is the target illuminant (R, G, B), and $\mathbf{E} = (R_E, G_E, B_E)$ is the estimate, then the angular error is given by $\cos^{-1}((\mathbf{T} \bullet \mathbf{E})|\mathbf{T}|^{-1}|\mathbf{E}|^{-1})$. The second error measure is the vector distance in (r, g) space of the illuminant chromaticity and the estimate thereof. Specifically, if $(r_T, g_T) = (R_T/S_T, G_T/S_T)$ where $S_T = R_T + G_T + B_T$ is the target illuminant chromaticity and $(r_E, g_E) = (R_E/S_E, G_E/S_E)$ where $S_E = R_E + G_E + B_E$ is the estimate, this second error is given by $\sqrt{(r_E - r_T)^2 + (g_E - g_T)^2}$. These two measures are roughly interchangeable, but the first makes more sense for the (R, G, B) algorithms, whereas the second is closer to the quantity that several of the chromaticity algorithms try to minimize.

To measure illuminant (R, G, B) error it is perhaps most natural to consider the vector distance in (R, G, B) between the illuminant and the estimate thereof, $((R_E - R_T)^2 + (G_E - G_T)^2 + (B_E - B_T)^2)^{1/2}$, and thus, we include some results using this error measurement. However, given the application dependent asymmetry between illuminant chromaticity and illuminant brightness, we find it more useful to look at the conjunction of a chromaticity error measure and the error in brightness. For the latter we use a fourth error measure, that being the difference in $R+G+B$ between the illuminant and the estimate thereof, $|S_E - S_T|$.

Finally, for our last two error measures we consider the error in the final color constancy result, which is the difference between the corrected image, and the exact target image taken under the canonical illuminant. These results are difficult to obtain with image data, because they require registered images with exactly the same geometry for each illuminant. This is only feasible if all illuminants are produced from a single source in conjunction with filters, which precludes the use of a general illuminant set like ours. The problems are diminished in the restricted case of chromaticity mappings, but even here, variations in geometry cause problems. With synthetic data it easy and useful to look at mapping results, and we use the RMS error over synthetic scene surfaces in (R, G, B) and (r, g) between the target data and the mapped estimate thereof. Specifically, if $\rho_i^{(k)}$ is the observed response for channel or chromaticity component k for pixel i, and $\tau_i^{(k)}$ is the analogous quantity for the target image, then the RMS error is computed by

$$\left(\frac{1}{N}\sum_{i}^{N}\frac{1}{K}\sum_{k}^{K}\left(\rho_i^{(k)} - \tau_i^{(k)}\right)^2\right)^{1/2}$$

where K is the number of channels (usually 3) or chromaticity components (usually 2) and N is the number of synthetic surfaces (or image pixels).

IV. Assumptions and Context

Most color constancy algorithms assume that the world consists of perfectly diffuse reflecting surfaces. Color constancy has also been attempted by making use of specularities [19]–[24]. In this paper we do not test any algorithms which specifically require specularities to be present. However, specularities are very common, and therefore we have study their effect on the various algorithms. We further assume that the illumination is the same for the entire (synthetic) scene.

Color constancy algorithms also generally make assumptions about the diversity, and possibly the statistics, of the surfaces and the illuminants that will be encountered. Typically the surfaces and illuminants are supplied as collections of surface

[1]The world (surfaces and illuminants) encountered by the camera also affects the diagonal model error.

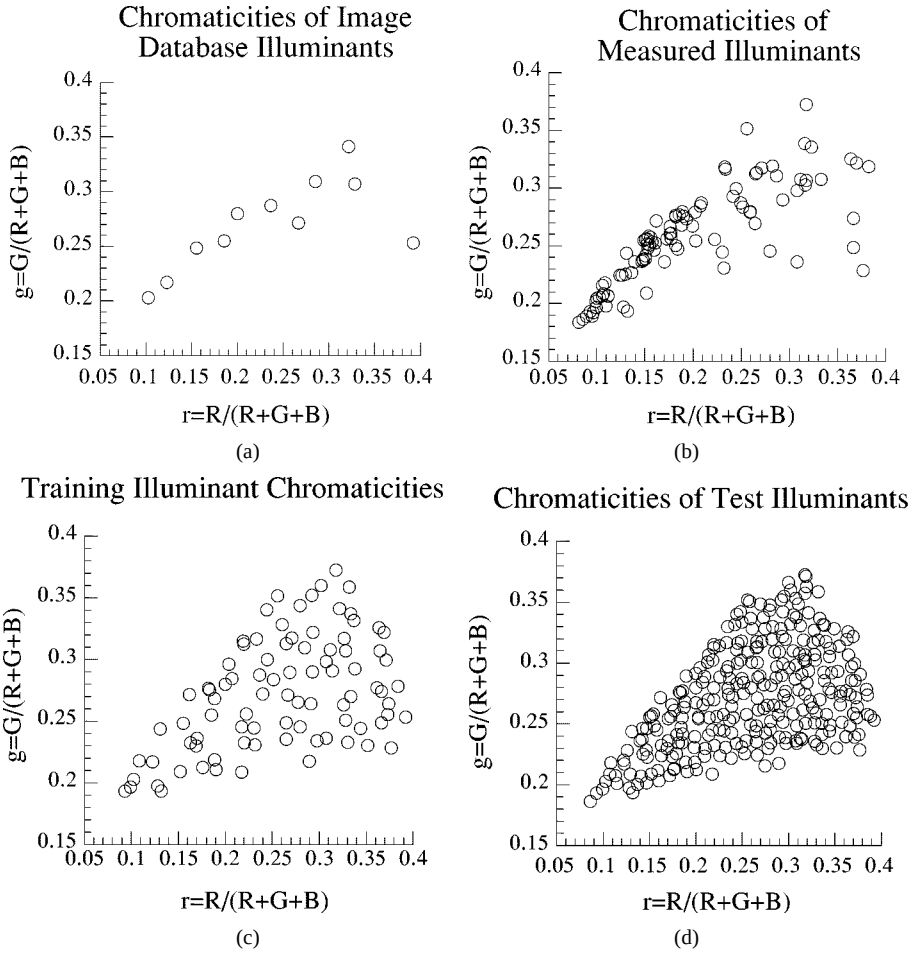

Fig. 1. Chromaticity distributions of the various sets of illuminants used in this study. The 11 illuminants used for creating test images are shown in (a). In (b), we plot the chromaticities of an additional set composed of more sources, including a number illuminations measured in and around our university campus. The training set constructed from these sources is shown in (c). A similar set used for testing with the chromaticity space more densely populated is shown in (d).

reflectances and illuminant energy spectra. The required data sets are then computed using an appropriate camera model.

For surface reflectances we used a set of 1995 spectra compiled from several sources. These surfaces included the 24 Macbeth color checker patches, 1269 Munsell chips, 120 Dupont paint chips [25], 170 natural objects [25], the 350 surfaces in the Krinov data set [26], and 57 additional surfaces measured by ourselves. This set was chosen to be a superset of the reflectance sets used by others for color constancy research. This set was used both for modeling the world for algorithm calibration (training) and testing the algorithms.

The illuminant spectra for all parts of this study were chosen to roughly uniformly cover the (r, g) chromaticities of common illumination conditions. All illuminant spectra were normalized so that a perfect white seen by our camera under each illuminant would have a maximum response among the three channels of 255. To obtain the appropriate illuminant sets, we first selected 11 sources to be used for image data experiments. These were selected to span the range of chromaticities of common natural and man made illuminants as best as possible, while bearing in mind the other considerations of stability over time, spectral nearness to common illuminants, and physical suitability for our experimental setup (using the sun would have been difficult). These 11 sources include three fluorescent lights (Sylvania warm white, Sylvania cool white, and Philips Ultralume), four different 12 volt incandescent lights, and those four used in conjunction with a blue filter (Roscolux 3202). The spectra of one of the incandescent lights (Sylvania 50MR16Q) is very similar to a regular incandescent light bulb. The other three are bulbs developed to provide spectra similar to daylight of three different color temperatures (Solux 3500K, Solux 4100K, Solux 4700K). When used in conjunction with the blue filter these three bulbs provide a reasonable coverage of the range of outdoor illumination. The chromaticities of all 11 illuminants are shown in Fig. 1(a).

To create the illuminant set used for training, we divided (r, g) space into cells 0.02 units wide, and placed the 11 illuminants described above into the appropriate cells. We then added illumination spectra from a second set of 97, provided that their chromaticity bins were not yet occupied. This second set consisted of additional sources and a number of illumination spectra measured in and around our university campus. The chromaticities of this additional illuminant set are shown in Fig. 1(b). Finally, to obtain the desired density of coverage, we used random linear combinations of spectra from the two sets. This is justified because illumination is often the blending of light from two or more sources. In addition, to the extent that the diagonal model holds, these constructed illumination spectra will behave like physical sources with the same chromaticities as the constructed ones. Fig. 1(c) shows the

chromaticities of the training set obtained using this method. Finally, to produce the illuminant set for testing, we followed the same procedure, but filled the space 4 times more densely. The resultant chromaticities are shown in Fig. 1(d).

We characterized our Sony DXC-930 CCD camera as described in [27] and used these sensors for generating camera responses for synthesizing scenes and algorithm training/calibration. The ideal (linearized) camera response for channel k, $\rho^{(k)}$, is computed from a surface spectra $S(\lambda)$ and illuminant spectra $E(\lambda)$ and sensor function $R^{(k)}$ by

$$\rho^{(k)} = \int R^{(k)}(\lambda)S(\lambda)E(\lambda)\,d\lambda. \tag{1}$$

In practice, all the functions are replaced by vectors. In our case we use 101 samples from 380 nm to 780 nm in steps of 4 nm which is the sampling provided by our PhotoResearch PR-650 spectrometer. The camera sensors, as well as the data used to estimate them, are available on-line [2].

Where relevant we use the Sylvania 50MR16Q for the canonical illuminant, as this is the illuminant for which the camera is best balanced. Specifically, under this illuminant, the camera response to perfect white is roughly the same across the three channels. The (r, g) chromaticities of the reflectance data set as imaged under the canonical illuminant and with the Sony DXC-930 camera are shown in Fig. 2.

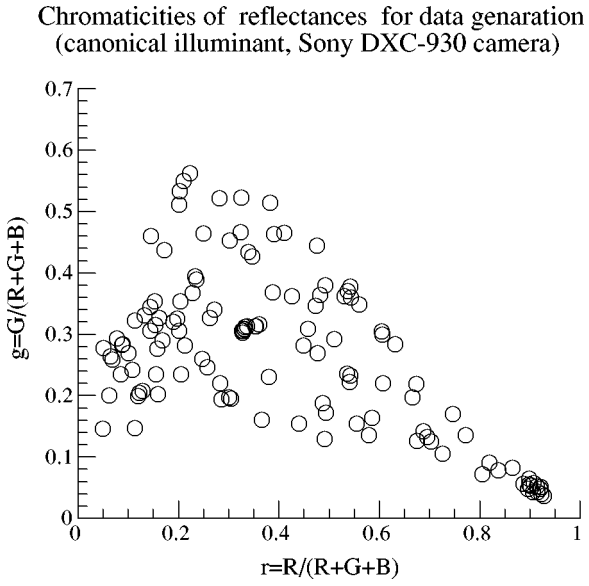

Fig. 2. Chromaticities of the reflectance data set as imaged under the canonical illuminant and with the Sony DXC-930 camera.

V. Algorithms

We endeavored to include the most promising computation color constancy algorithms, as well as several simple, yet effective, commonly used algorithms. Several algorithms that were not implemented bear mentioning. First, we did not test the innovative Maloney–Wandell algorithm. Despite being an important contribution to the development of many ideas, this algorithm simply does not work well in the general context in which we test color constancy. The reason for this is that, for a three sensor vision system, this algorithm requires that the surface reflectances of the world can be well approximated by two basis functions. This is not true in general. Several authors have noted that this problem leads to poor performance [28], [29].

A second important algorithm not tested is Brainard and Freeman's Bayesian method [28]. Instead we investigate the related Color by Correlation method as originally introduced for chromaticity input [16], [30], [31] (Brainard and Freeman's approach [28] use a three-dimensional (3-D) space; a 3-D version of Color by Correlation has also recently been proposed [32]; also see [33] for another approach related to Color by Correlation).

We also exclude Buchsbaum's gray world variant [6], as well as that of Gershon *et al.* [7] which use linear models in conjunction with gray world like assumptions. An additional part of Gershon *et al.*'s algorithm is the idea that gray world averaging should be done over segmented components of an image, rather than the image pixels. This is implicit when synthesized data is used. With image data this strategy can be implemented with appropriate pre-processing.

Finally, the extensive body of work on the Retinex theory of human vision has yielded several algorithms. The emphasis of Retinex theory is on human vision, and goes beyond simple illuminant estimation. Hence, computational color constancy algorithms emerge from Retinex more as a process of analogy than through specification by the original researchers. Nonetheless, at least three algorithms for simple illuminant estimation in our context can be identified. We investigate the one closest to [34]–[36]. We do not test the method in [37], [38] (analyzed in [39]), nor the method in [40], as they are essentially gray world algorithms. We now discuss the algorithms chosen for detailed study.

A. Gray World Methods

The gray world method assumes that the average of the surface reflectances of a typical scene is some pre-specified value, which is referred to as "gray." The exact definition of "gray" requires some clarification. One possibility is simply true gray; specifically, a 50% uniform reflectance. This leads to the algorithm labeled GW in the results. A second choice is to use the average of the reflectance database. This is expected to perform better then GW with synthesized data, and the result is guaranteed to be excellent if a large number of surfaces are used. With image data, however, the actual average surface reflectance is not known *a priori*, and thus, this method is expected to fare relatively less well. We denote this algorithm by DB-GW. We note that since these algorithms work on camera sensor responses, the actual assumption about scene averages is weaker than stated above. Specifically, the algorithms simply assume that the scene average is identical to the camera response to the chosen "gray" under the scene illuminant. Then, under the diagonal assumption, the color of white can be estimated from that average. In the case of GW, the average is simply multiplied by two. In the case of DB-GW, we scale the result by the ratio of the camera response to white under the canonical illuminant, to the camera response to gray, again under the canonical illuminant.

B. Illuminant Estimation by the Maximum of Each Channel

The SCALE-BY-MAX algorithm estimates the illuminant (R, G, B) by the maximum response in each channel. It is a limiting case of one version of Retinex [34], [35], [39], [41]. This method is clearly sensitive to the dynamic range of the vision system. We also note that for a world of matte reflectances, the estimate of the illuminant magnitude provided will be biased, as the maximum reflectance in the scene will always be less than that of a pure white. It would seem feasible to compensate for this bias, but it would be a significant undertaking to accurately estimate it, as it changes with the number of surfaces in the scene. Also, if specularities are present, then the maximum reflectance can easily be greater than that of pure white. On the positive side, we note that if significant specularities are present, and the vision system has sufficient dynamic range to prevent them from being clipped, then this method provides an excellent estimate of illuminant chromaticity. In doing so, the algorithm is implicitly making use of pixel brightness information and thus, can potentially out-perform algorithms which use only chromaticity input.

C. Gamut Mapping Methods

We present the results of a number of algorithms based on Forsyth's gamut-mapping approach [8] (see also [1], [12], [15], [17], [42]). The first step of the approach is to form the set of all possible (R, G, B) due to surfaces in the world under a known, "canonical" illuminant. This set is convex and we represent it by its convex hull. Similarly, we represent the set of all possible (R, G, B) under the unknown illuminant by its (unknown) convex hull. Under the diagonal assumption of illumination change, these two hulls are a unique diagonal mapping (a simple 3-D stretch) of each other. The goal is to estimate that diagonal mapping.

Fig. 3 illustrates the situation using triangles to represent the gamuts. In the full (R, G, B) version of the algorithm, the gamuts are actually 3-D polytopes. The upper thicker triangle represents the unknown gamut of the possible sensor responses under the unknown illuminant, and the lower thicker triangle represents the known gamut of sensor responses under the canonical illuminant. We seek the mapping between the sets, but since the one set is not known, we estimate it by the observed sensor responses. These responses form a subset, the convex hull of which is illustrated by the thinner triangle. Because the observed set is normally a proper subset, the mapping to the canonical is not unique, and Forsyth [8] provides a method for effectively computing the set of possible diagonal maps which is a convex set in the space of mapping coefficients.

Since each observed (R, G, B) must be mapped into the canonical gamut, a candidate diagonal map (d_R, d_G, d_B) must satisfy $(Rd_R, Gd_G, Bd_B) \in \boldsymbol{C}$, where $\boldsymbol{C}$ is the canonical gamut. This means in turn that $(d_R, d_G, d_B) \in \boldsymbol{C}_{(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})}$, were $\boldsymbol{C}_{(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})}$ is obtained from $\boldsymbol{C}$ by dividing each (hull) point by (R, G, B) element-wise. Each such set is a (convex) constraint on the possible diagonal maps, and the final solution set is determined by intersecting the collection of convex sets obtained by considering each observed (R, G, B)—in fact, it suffices to intersect only the sets corresponding to the vertex points of the convex hull of the observed (R, G, B). To find convex hulls we use the freely available program "qhull" [43]. Convex sets can intersected by breaking space into cubes and collecting those which are in all hulls, or much more elegantly by taking the convex hull of the appropriate quantities in dual space [44].

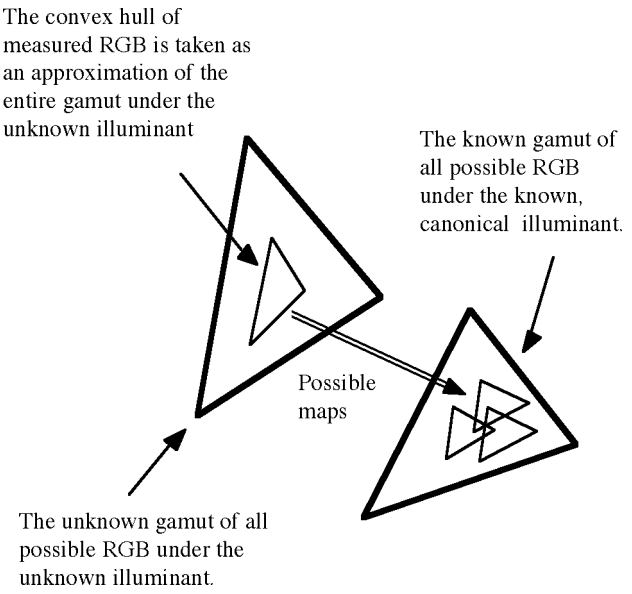

Fig. 3. Illustration of the basic idea of gamut-mapping color constancy.

It is possible that the sets to be intersected have no point in common as a result of failures in the assumptions or other errors such as noise. Here we augment the observed data with the corner points of error boxes around the data, and compute contributions to $\boldsymbol{C}_{(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})}$ for each corner point. The modified $\boldsymbol{C}_{(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})}$ is again the convex hull of the contributing maps, and is expanded as a result of modeling the error. Note that this is not that same as simply mapping an expanded observed hull, as the observed points need to be inverted to find $\boldsymbol{C}_{(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})}$. We threshold the error box corner points so that all R, G, or B used are larger than a small positive value to avoid problems when they are inverted. The amount of error is slowly increased until the intersection is nonempty.

Finlayson's Color in Perspective algorithm [15] adds two additional ideas to the gamut mapping method. First, the method can be used with the chromaticity space $(R/B, G/B)$. Second, the diagonal maps can be further constrained by restricting them to ones corresponding to expected illuminants. This new constraint is nonconvex in the space of diagonal maps. In [15] the combined solution set was considered to be the intersection of the convex constraint set due to the original surface constraints, and the nonconvex illuminant constraint set. In [42] the illuminant constraint set was approximated by its convex hull and also used in the full (R, G, B) case.

Once the set of possible maps has been computed, an important second stage of the algorithm is to choose a solution from the feasible set. Several different methods for doing this have been proposed. The original method chose the solution which maximized the volume of the mapped set [8], which is simply the diagonal transform with maximal determinant.

The Color in Perspective method uses the same heuristic in chromaticity space. However, this solution method is quite biased, and in [42] the average of the constraint set was investigated, both in the chromaticity based algorithm and the (R, G, B) algorithm. This method for choosing the solution is still biased in the chromaticity case, and in [17] the averaging was done in three dimensions. Specifically, the constraints on the mappings in perspective space correspond to cones in the space of mappings between (R, G, B) gamuts. In order to average over the nonconvex illumination constraint, Monte Carlo integration was used. In this work, we approximate this average using a more direct form of numerical integration, both in the chromaticity and (R, G, B) cases. We simply divide space into cubes (or squares) and average those for which the implied illuminant is in the cone of possible illuminants. The implied illuminant is found by dividing the canonical illuminant (R, G, B) [or $(R/B, G/B)$] by the mapping corresponding to the chosen cube (or square) element-wise.

To summarize, we investigate three methods of forming the solution set. These are Forsyth's original method, designated by CRULE (for "coefficient-rule," the name of the original algorithm), the Color in Perspective method, designated by CIP, and the illumination constraint set applied to CRULE designated by ECRULE (for "extended-CRULE"). We do not consider the chromaticity case without the illumination constraint. These solution sets are paired with three methods of selecting a solution from them. We use MV to denote the original maximum volume heuristic, AVE to specify that the constraint set is averaged, using a convex approximation to the illumination constraint if necessary, and ICA to specify that the constraint set is numerically integrated because it is nonconvex ("illumination constrained average").

For our experiments the canonical gamut is determined by first calculating all (R, G, B) of the reflectance set under the canonical illuminant (the Sylvania 50MR16Q), and then computing the convex hull of that set. The (R, G, B) of the canonical illuminant is simply the (R, G, B) of a pure white as imaged under that illuminant. We assume that the illuminant set is conical (all illuminants can be arbitrarily bright), and it is computed by projecting the (R, G, B) of the illuminant data set onto the (r, g) chromaticity plane $R+G+B = 1$, and taking the convex hull of the projected points. The illumination gamut is then the cone defined by the origin and the hull points on the plane. Rather than use this cone as a 3-D entity, we can often simply use it implicitly by observing that an (R, G, B) is inside the cone, if the corresponding (r, g) is in the convex hull of the projected points used to define the cone.

D. Color by Correlation

Recently, Finlayson *et al.* introduced Color by Correlation [16], [30], [31] as an improvement on the Color in Perspective method. The basic idea of Color by Correlation is to pre-compute a correlation matrix which describes the extent to which proposed illuminants are compatible with the occurrence of image chromaticities. Each row in the matrix corresponds to a different training illuminant. The matrix columns correspond to possible chromaticity ranges resulting from a discretization of (r, g) space, ordered in any convenient manner. Two versions of Color by Correlation are described in [16]. In the first version, the elements of the correlation matrix corresponding to a given illuminant are computed as follows: First, the (r, g) chromaticities of the reflectances in the training set under that illuminant are computed using the camera sensors. Then the convex hull of these chromaticities is found, and all chromaticity bins within the hull are identified as being compatible with the given illuminant. Finally, all entries in the row for the given illuminant corresponding to compatible chromaticities are set to one, and all other elements in that row are set to zero.

To estimate the illuminant chromaticity, the correlation matrix is multiplied by a vector whose elements correspond to the same (r, g) bins used in the correlation matrix. The elements of this vector are set to one if the corresponding chromaticity occurred in the image, and zero otherwise. The ith element of the resulting vector is then the number of chromaticities which are consistent with the illuminant. Under ideal circumstances, all chromaticities in the image will be consistent with the actual illuminant, and that illuminant will therefore have maximal correlation. As is the case with gamut-mapping methods, it is possible to have more than one plausible illuminant, and in our implementation we use the average of all candidates close to the maximum. This algorithm is quite close to Color in Perspective and provides a convenient alternative implementation. We label this algorithm "C-by-C-01."

In the second version of Color by Correlation, the correlation matrix is set up to compute the probability that the observed chromaticities are due to each of the training illuminants. The best illuminant can then be chosen using a maximum likelihood estimate, or using some other estimate as discussed below. To compute the correlation matrix, the set of (r, g) for each illuminant is again found using our database of surface reflectances. The frequency of occurrence of each discrete (r, g) is then recorded. If additional prior information about the probability of occurrence of these reflectances is available, then the counts are weighted accordingly. Illuminant priors are also supported. In our implementation we simply use uniform statistics. The constructed counts are proportional to the probability that a given (r, g) would be observed, given the specific illuminant. The logarithms of these probabilities for a given illuminant are stored in a corresponding row of the correlation matrix. The application of the correlation matrix, done exactly as in the C-by-C-01 case, now computes the logarithm of the posterior distribution.

This computation of the posterior distribution is a simple application of Bayes's rule. Specifically, the probability that the scene illuminant is I, given a collection of observed chromaticities C, is given by

$$P(I|C) = \frac{P(C|I)P(I)}{P(C)}. \tag{2}$$

Since we are assuming uniform priors for I, and since $P(C)$ is a normalization which is not of interest, this reduces to

$$P(I|C) \propto P(C|I). \tag{3}$$

Assuming that the observed chromaticities are independent, $P(C|I)$ itself is the product of the probabilities of observing the individual chromaticities c, given the illuminant I

$$P(C|I) = \prod_{c \in C} P(c|I). \tag{4}$$

Taking logarithms gives

$$\log(P(C|I)) = \sum_{c \in C} \log(P(c|I)). \tag{5}$$

This final quantity is exactly what is computed by the application of the correlation matrix to the vector of chromaticity occurrences. Specifically, the ith element of the resulting vector is the logarithm of the posterior probability for the ith illuminant.

There are several potential problems with the method as described so far. First, due to noise, and other sources of mismatches between the model and the real world, an observed set of chromaticities can yield zero probability for all illuminants, even if the illuminant, or a similar one, is in the training set. Second, the illumination may be a combination of two illuminants, such as an arbitrary mix of direct sunlight and blue sky, and ideally we would like the method to give an intermediate answer. We deal with these problems as follows. First, as described above, we ensure that our illuminant set covers (r, g) space, so that there is always a possible illuminant not too far from the actual illuminant. Second, as we build the correlation matrices, we smooth the frequency distribution of observed (r, g) with a Gaussian filter $(\sigma = 1.0)$. This ensures that there are no holes in the distribution, and compensates for noise.

The final step is to choose an answer, given the posterior probability distribution. The original work [16] suggests three choices: The maximum likelihood, mean likelihood, and the local area mean, introduced in [28]. That work discusses these methods in detail with respect to a related Bayesian approach to color constancy, where they are referred to as the MAP, MMSE, and MLM estimators, respectively. We will adopt this notation here as suffixes to "C-by-C." The MAP estimate is simply the illuminant which has the maximum posterior probability. To compute the MMSE estimate of the chromaticity estimate we take the average (r, g) weighted by the posterior distribution. The MLM estimator is computed by convolving the posterior distribution with a Gaussian mask, and then finding the maximum. For our purposes, we would like to choose the particular Gaussian mask sigma which minimizes the error of some specific task. Unfortunately, the bulk of our results are not of much help here, as they are based on RMS error, and thus, we already know that the MMSE method will work better. Brainard and Freeman argue that the MLM estimate should be considered in favor of the MMSE estimate because the latter penalizes large errors too much. They would prefer that once errors are beyond a certain size, they are considered equally bad. Without additional knowledge about the application, it is difficult to quantify such an error measure, but we can take a small step in that direction by considering the average absolute error, as opposed to the RMS error. This yields a sigma where the MLM estimate is slightly better than the MMSE estimate for average absolute error, and we use this sigma (8.0) for the MLM estimates reported. However, demonstrating the virtues of the MLM method would require error measures which are different than the ones used for this study.

E. Neural Net Methods

We also provide the results from a neural network trained to estimate the color of the illuminant [18], [45], [46] (Labeled NEURAL-NET in the results). The neural net is a multilayer Perceptron with two hidden layers. As is common, the general structure is pyramidal. The input layer consists of 2500 nodes, the first hidden layer has 400 nodes, the second hidden layer has 30 nodes, and the output layer has 2 nodes. We divide (r, g) chromaticity space into discrete bins, with each input neuron corresponding to one of the discrete bins. The input to each neuron is a binary value representing the presence or absence of a scene chromaticity falling in the corresponding (r, g) bin. Thus, we form a (r, g) histogram of the image, and then binarize that histogram.

The output signal from the two output neurons are real valued, and correspond to an estimate of the chromaticity of the scene illuminant. Output signals are computed as a weighted sum of values of input neurons put through a sigmoid function. The network is trained to compute this estimate by being presented with many synthesized images, generated from the training sets described above, together with the chromaticity of the illuminant used to generate each image. The training of the neural net occurs by re-adjustment of neuron weights using back-propagation without momentum, based on the discrepancy between predicted and actual scene illuminant chromaticity. Extensive details are provided in [45], [46].

VI. Experiments

In each of three experiments we computed the performance of the algorithms for synthetic scenes with 4, 8, 16, 32, 65, 128, 256, 512, and 1024 surfaces. For each number of surfaces, we generated 1000 scenes with the surfaces randomly selected from the reflectance database and a randomly selected illuminant from the test illuminant database. For each error measure, algorithm, and number of scenes we computed the root mean square error (RMS) over the 1000 results. More specifically, for a given error measure, let E_i be the error for the ith synthesized scene. Then the RMS error is given by

$$\left(\frac{1}{N} \sum_{i}^{N} E_i^2 \right)^{1/2}$$

where N is 1000 in this case. We chose RMS over the average because, on the assumption of roughly normally distributed errors with mean zero (approximately true for most algorithms [1, pp. 82–86]), the RMS gives us an estimate of the standard deviation of the algorithm estimates around the target. This is preferable to using the average of the magnitude of the errors, as those values are not normally distributed. Finally, given normal statistics, we can estimate the relative error in the RMS estimate by $1/\sqrt{2N}$ [47, p. 269]. For $N = 1000$, this is roughly 2%.

For each experiment we consider the results for eight surfaces in detail. This number of surfaces is closest in difficulty to the data from a comprehensive set of images taken with a

real camera [1], [4]. Although absolute errors found with synthesized and captured data are not generally comparable, we are interested in studying the changes in relative performance over the two conditions. This is most valid if the numbers are roughly the same, and we ensure this by specifying eight synthetic surfaces for comparison with our image data results.

We include the results of two minimal color constancy methods. The first (NOTHING) is to do nothing, which implicitly assumes that the vision system is already properly calibrated for the actual illuminant, and, in our context, is equivalent to guessing that the actual illuminant is the canonical (target) illuminant. The second method (AVE-ILLUM) is similar, but instead, the illuminant is assumed to be the average of the normalized illuminants in our database.

A. Experiment With Matte Data

In Fig. 4, we plot the error in (r, g) for selected algorithms as a function of the number of matte surfaces. As the number of surfaces in the scenes increases, there is more information available for the algorithms, and performance generally increases. For a sense of the absolute scale, we offer the heuristic that an error of 0.02 is adequate color constancy for most tasks; the current state of the art falls short of this mark. In Table I, we provide the results of all algorithms using several error measures for eight surfaces.

Both NOTHING and AVE-ILLUM are independent of the scene, and thus, their error is constant with respect to the number of surfaces. Since the test illuminants are distributed throughout the data set, and since the canonical illuminant is toward the periphery of the set (it is redder than average), AVE-ILLUM is a more effective minimal algorithm than NOTHING. A similar consideration explains the poor performance of the CIP-MV algorithm. As found in [1], this algorithm is biased. Specifically, in the $(R/B, G/B)$ chromaticity space, the maximum volume constraint chooses essentially the bluest illuminant consistent with the observed chromaticities. With our comprehensive illuminant data set, many surfaces are required before the algorithm CIP-MV performs better than the two minimal algorithms. Similarly, many surfaces are also required to obtain a good result using the somewhat less biased CIP-HA algorithm. Finally the CIP-ICA algorithm was consistently better than both NOTHING and AVE-ILLUM, and performed much better than the other two Color in Perspective methods. Also, as expected, the C-by-C-01 algorithm behaved similarly to the CIP-ICA method.

The rank order of algorithm performance is a function of the number of surfaces. For example, the CIP-ICA and C-by-C-01 methods performed better than SCALE-BY-MAX for a small number of surfaces, but the error with SCALE-BY-MAX dropped rapidly as the number of surfaces increased, becoming lower than that for the Color in Perspective methods at around eight surfaces. The methods based on 3-D gamut-mapping generally did better than SCALE-BY-MAX, although the variants which average their constraint set to obtain the final solution (CRULE-AVE, ECRULE-AVE, and ECRULE-ICA) were exceeded by SCALE-BY-MAX at around 16 surfaces. The maximum volume heuristic was better than averaging for estimating illuminant chromaticity, except in the important case of a small number of surfaces, where the ECRULE-ICA method was the best gamut-mapping algorithm.

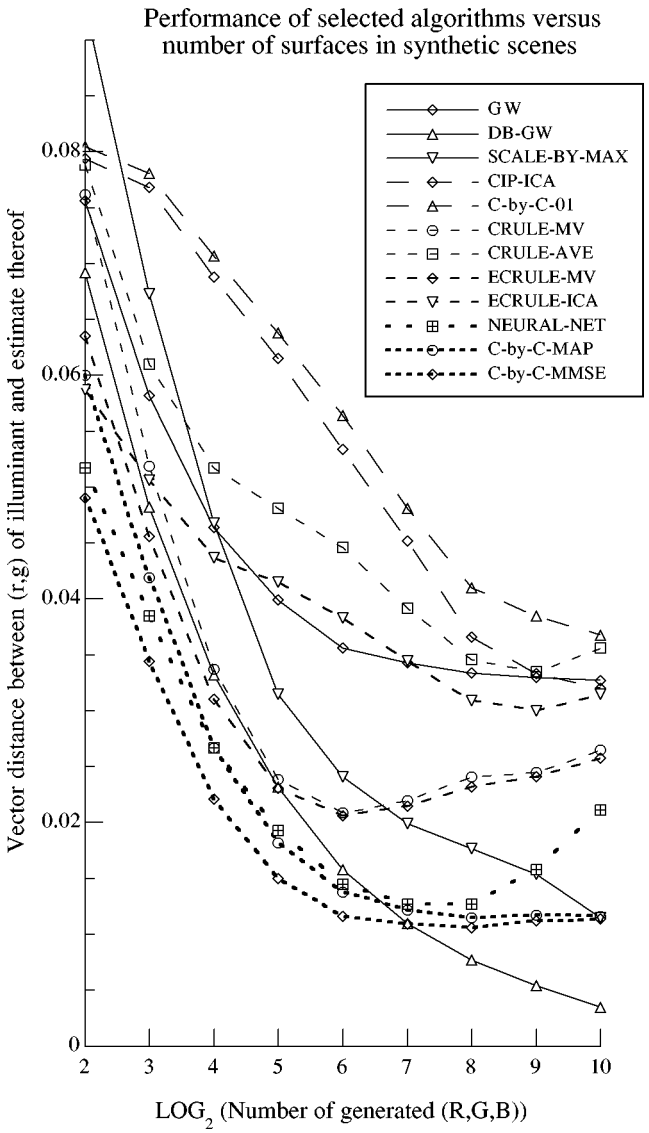

Fig. 4. Error in (r, g) chromaticity as a function of the number of surfaces for some of the better performing algorithms. The results of NOTHING, AVE-ILLUM, CIP-MV, and CIP-HA are largely off the scale, and, thus, are omitted. Numeric results for the case of eight surfaces (corresponding to $x = 3$ in this plot) for all algorithms are available in Table I.

In general, the methods which make use of the distribution of the input chromaticities, specifically Color by Correlation and the neural net, gave the best chromaticity estimates. As expected, the C-by-C-MMSE algorithm performed better than the C-by-C-MAP algorithm, as it is known to be optimal for RMS error. The error for the neural net method was between C-by-C-MAP and C-by-C-MMSE, as was C-by-C-MLM. We emphasize that these results are for the particular statistical environment we chose for testing. In fact, the environment we chose is difficult in two regards: First, we use a relatively large range of illuminants, especially in the direction away from the daylight illuminant locus. Our choice is based on illumination environments which we have measured in and around our university campus, but for some applications it is quite possible that the illuminant set could be significantly restricted, or suitable

TABLE I
ALGORITHM PERFORMANCE WITH RESPECT TO A NUMBER OF ERROR MEASURES. THE VALUES LISTED ARE THE RMS OVER 1000 SYNTHETICALLY GENERATED SCENES, EACH HAVING EIGHT SURFACES RANDOMLY SELECTED FROM THE REFLECTANCE DATA SET, AS VIEWED UNDER AN ILLUMINANT RANDOMLY SELECTED FROM THE TEST ILLUMINANT DATA SET. THE UNCERTAINTY IN THESE NUMBERS IS ROUGHLY 2%. AN ASTERISK IS USED FOR MISSING OR NONAPPLICABLE VALUES

Algorithm	Illuminant Estimate Angular Error (degrees)	Illuminant Estimate rg Error	Illuminant Estimate RGB Error	Illuminant Estimate R+G+B Error	Scene Mapping rg error (RMS over pixels)	Scene Mapping RGB error (RMS over pixels)
NOTHING	16.45	0.114	*	*	0.113	*
AVE-ILLUM	11.79	0.086	*	*	0.089	*
GW	8.00	0.058	190	310.1	0.062	137
DB-GW	6.51	0.048	96	144.6	0.054	33
SCALE-BY-MAX	9.03	0.067	165	267.0	0.072	106
CIP-MV	26.27	0.200	*	*	0.240	*
CIP-AVE	18.12	0.130	*	*	0.141	*
CIP-ICA	10.51	0.077	*	*	0.081	*
NEURAL-NET	5.23	0.038	*	*	0.045	*
C-by-C-01	10.79	0.078	*	*	0.082	*
C-by-C-MAP	5.63	0.042	*	*	0.048	*
C-by-C-MLM	5.25	0.039	*	*	0.045	*
C-by-C-MMSE	4.66	0.034	*	*	0.041	*
CRULE-MV	6.75	0.052	111	178	0.058	54
CRULE-AVE	8.39	0.061	208	303	0.066	45
ECRULE-MV	6.04	0.046	107	172	0.052	51
ECRULE-AVE	7.22	0.051	147	223	0.054	37
ECRULE-ICA	7.15	0.051	144	218	0.053	37

priors could be placed on the illuminants, and this would help the algorithms which make use of prior knowledge about the illuminant. In this work we have imposed a relatively uniform prior on the illuminant (r, g), which is difficult to exploit—for a typical error measure, averaging of feasible solutions will do just as well.

Second, we expect our reflectance data is also relatively difficult for the methods which attempt to exploit detailed statistics of the world. Although it is not uniform in (r, g) space (which would be even more difficult), the over-riding characteristic, as exemplified by the Munsell data (a significant subset), is that it covers the range of colors reasonably well. If we had first generated data with a more peaked distribution, then we would expect that both the neural net and Color by Correlation would do even better. However, a model of what to expect in a random image is not readily available, and thus, we chose to investigate the algorithms in the context of no preference for any given surface beyond that implied by our data set. This means that the results presented here are likely to be closer to the worst rather than best case for these algorithms. In fact we expect that some of the excellent results reported using Color by Correlation [30] are due to narrower assumptions regarding the statistics of illuminants and surface reflectances. Unfortunately, the details of the algorithm as it is used in [30] are not available due to commercial interests, and therefore it is difficult to provide further analysis.

We note that for a large number of synthetic surfaces, which is much better input than normally available in practice, the errors of most algorithms did not converge exactly to zero. In the case of the Color by Correlation methods, this is simply due to the discretization of the input. We chose to break the (r, g) space into 50 units by 50 units, which corresponds to bins which are 0.02 units square. Thus, we cannot expect the algorithm to do much better than an error of 0.01 in each of r and g, which corresponds to a vector distance error of $0.01 * \sqrt{2}$, which is consistent with our findings. We chose the resolution to be consistent with that of the neural network, and we assume that the limiting error for the neural net is also largely due to the same problem. This error could be reduced by using a more accurate discretization, but doing so is not particularly important, as we rarely have this kind of input outside of simulation.

The gamut-mapping methods also did not converge to zero error. Here the problem is the failure of the diagonal model which relates the solution set (in the space of diagonal maps) to illumination estimates. The error for SCALE-BY-MAX, on the other hand, did get close to zero. It does not go exactly to zero because we did not include a perfect reflectance in the data set. The GW algorithm converged to a specifiable error, due to the difference between the actual database average, and a perfect gray. Finally, when we used the database average for gray with the DB-GW algorithm, the error converged to zero as expected.

We turn briefly to brightness. In general, none of the algorithms provide good estimates of illuminant $R + G + B$ with eight surfaces. The lowest error is 145 (DB-GW) which is large given the range (0, 765). The superior performance of the DB-GW algorithm is due to its calibration on the true mean reflectance of the surfaces. Unfortunately, this statistic is not typically available *a priori* for image data. Among the other algorithms, the best choice is invariably one of the gamut-mapping algorithms. We make the observation that the preferred gamut mapping algorithm depends on the error measure. If the goal is to estimate the illuminant (R, G, B) or $R + G + B$ then we prefer the maximum volume heuristic. On the other hand, if we wish to accurately map the image to a similar one of the same scene under the canonical illuminant [(R, G, B) mapping error—rightmost column in Table I], then we prefer choosing the solution from the constraint set by averaging. This is consistent with the easily demonstrated

claim that, under uniform statistics, the average minimizes the expected RMS mapping error [42], which was the reason for introducing this solution selection method.

Given that the illuminant (R, G, B) and the selected transform are related through the diagonal model, the observation that the preferred solution selection method depends on the error measure requires further explanation. The important observation is that a given diagonal map is approximately proportional to the element-wise *inverse* of the illuminant (R, G, B). This is most easily visualized by noting that as we approach the origin in the mapping space, we are dealing with an increasingly bright illuminant. This inversion relating the mapping space to the illuminant space means that averaging in one space (diagonal maps) is not equivalent to averaging in the other (illuminants). The inversion also means that an assumption of uniform statistics in the one space does not mean uniform statistics in the other. Thus, we do not necessarily expect the average of the diagonal maps to give the best illuminant (R, G, B) estimate.

B. Experiment With Simulated Specularities

In a second experiment, we simulated scenes where each surface was made specular with a probability of 25%. To each of these reflectance spectra, we added a random factor times a perfect reflectance. The random factors were uniformly distributed between zero and two. Thus, the surface reflectances could now be up to three times as bright as in the previous experiment.

We plot the change in the chromaticity results from the nonspecular case as a function of the number of surfaces (Fig. 5), and provide a variety of absolute results for the case of eight surfaces (Table II). Here we see that when specularities are present, most algorithms estimate chromaticity more accurately (but with widely differing improvements), even though they were not designed to take advantage of specularities. For example, with specular reflection, the maximum value in each channel is more likely to be close to the color of white under the scene illuminant, and thus, in this test, SCALE-BY-MAX does especially well.

Chromaticity based algorithms cannot make use of the brightness information, but as argued in [15] specularities should not have an overly negative impact because they simply move observed chromaticities toward white. These modified chromaticities are valid for some physically realizable surface, and thus, basic assumptions are not violated. However, we do expect some degradation given our experimental paradigm, because colors which are more saturated tend to lead to stronger constraints on the illuminant. This is consistent with the drop in the performance of the Color in Perspective method, clear in Fig. 5.

We also expected gray world algorithms to estimate the illuminant chromaticity better when specularities are present because the extra part added to the sum used in the average has the same chromaticity as a perfect gray. We note that the extra part does not have the same chromaticity as the gray used by the DB-GW algorithm, which explains why that algorithm degrades in the case of a large number of surfaces.

Amongst the 3-D gamut-mapping algorithms, the largest performance increase due to specularities was when the maximum volume heuristic was used to choose the solution. This was expected, as this heuristic tends to choose the map which takes a bright (and thus, in this experiment, specular) pixel to the

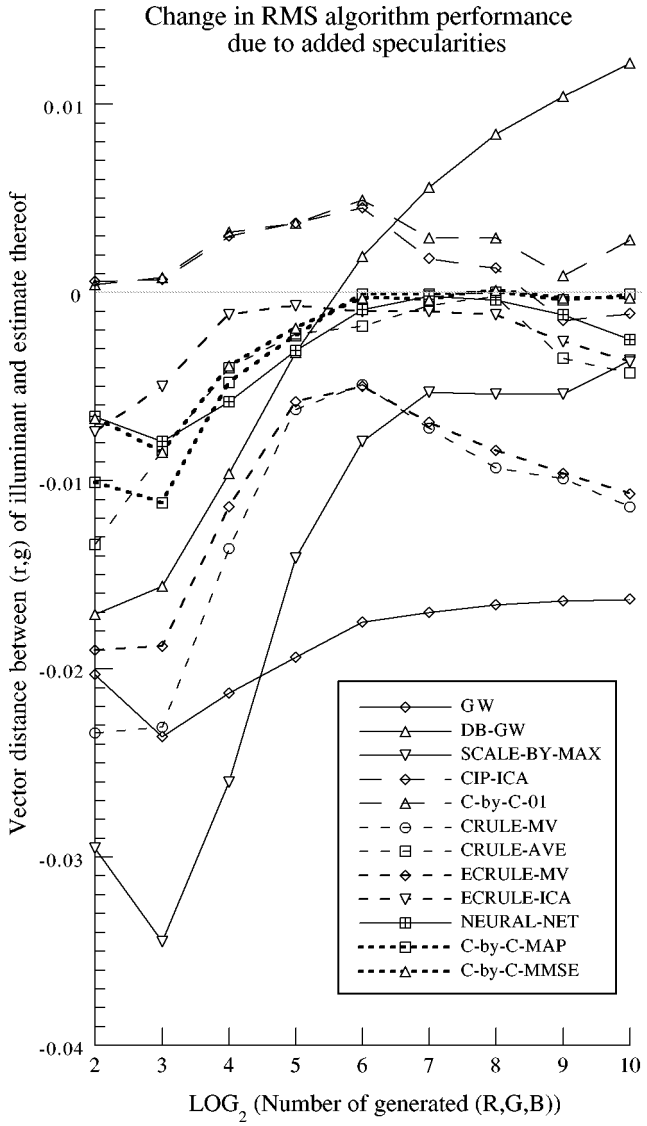

Fig. 5. Change in (r, g) error due to the addition of simulated specularities. The experiment used to produce Fig. 3 was rerun with added specularities. This plot is the difference of the errors, as a function of the number of surfaces. The absolute errors for the case of eight surfaces (corresponding to $x = 3$ in this plot) are available in Table III.

(R, G, B) for white under the canonical, which yields a good chromaticity result.

The results in Table II also show that specular reflection has a generally negative impact on the estimation of the overall brightness of the illuminant. This is expected, as none of the algorithms model the effect of specular reflection on pixel brightness. The decrease in error in the GW case is simply due to a brightness bias in that algorithm in the nonspecular case, and thus, it essentially works better by accident.

C. Experiment With Simulated Specularities and Clipping

When strong specularities are present in real images, they are often clipped, which can significantly affect algorithm performance. This motivated our third experiment where we investigated the performance of the algorithms in the specular case with simulated clipping of the brightest pixels. Three levels of artificial clipping were used. For each level, all pixels with R,

TABLE II
ALGORITHM PERFORMANCE WITH RESPECT TO A NUMBER OF ERROR MEASURES FOR SYNTHETIC SCENES (8 SURFACES) WITH SPECULARITIES (UNCERTAINTY IS ROUGHLY 2%). A SYNTHETIC SPECULARITY WAS ADDED TO 25% OF THE SURFACES USED BY ADDING A RANDOM FACTOR, UNIFORMLY DISTRIBUTED BETWEEN 1 AND 2, OF A PERFECT REFLECTANCE. OTHERWISE THESE RESULTS ARE ANALOGOUS TO THAT IN TABLE I

Algorithm	Illuminant Estimate Angular Error (degrees)	Illuminant Estimate rg Error	Illuminant Estimate RGB Error	Illuminant Estimate R+G+B Error	Scene Mapping RMS rg error	Scene Mapping RMS RGB Error
NOTHING	16.4	0.114	*	*	0.114	*
AVE-ILLUM	11.8	0.086	*	*	0.090	*
GW	4.8	0.035	132	216.5	0.041	104
DB-GW	4.5	0.033	448	731.4	0.038	170
SCALE-BY-MAX	4.5	0.033	257	427.8	0.039	140
CIP-MV	26.5	0.202	*	*	0.245	*
CIP-AVE	18.3	0.131	*	*	0.143	*
CIP-ICA	10.6	0.078	*	*	0.082	*
NEURAL-NET	4.1	0.031	*	*	0.037	*
C-by-C-01	10.9	0.079	*	*	0.083	*
C-by-C-MAP	4.2	0.031	*	*	0.037	*
C-by-C-MLM	3.7	0.027	*	*	0.034	*
C-by-C-MMSE	3.5	0.026	*	*	0.032	*
CRULE-MV	3.8	0.029	349	577	0.035	156
CRULE-AVE	7.5	0.053	1027	1632	0.054	219
ECRULE-MV	3.6	0.027	350	579	0.033	156
ECRULE-AVE	6.6	0.046	825	1353	0.049	210
ECRULE-ICA	6.5	0.046	811	1331	0.048	209

TABLE III
ALGORITHM CHROMATICITY PERFORMANCE WITH RESPECT TO TWO ERROR MEASURES FOR GENERATED SCENES WITH SPECULARITIES WITH THREE LEVELS OF SIMULATED CLIPPING (UNCERTAINTY IS ROUGHLY 2%). ALL GENERATED (R, G, B) VALUES WITH ANY OF R, G, OR B OVER THE CLIPPING LEVEL WERE DISCARDED. IN ALL OTHER REGARDS, THE EXPERIMENT WHICH PRODUCED THE NUMBER HERE IS ANALOGOUS TO THAT FOR TABLE II

	Illuminant Estimate Angular Error (degrees)		
Algorithm	Clipped at 400	Clipped at 300	Clipped at 250
NOTHING	16.4	16.4	16.4
AVE-ILLUM	11.8	11.8	11.8
GW	6.2	7.3	7.8
DB-GW	5.4	6.2	6.6
SCALE-BY-MAX	6.3	7.9	8.7
CIP-MV	26.5	26.5	26.5
CIP-AVE	18.3	18.3	18.3
CIP-ICA	10.6	10.6	10.6
NEURAL-NET	4.6	5.1	5.4
C-by-C-01	10.9	10.9	10.9
C-by-C-MAP	4.7	5.5	5.8
C-by-C-MLM	4.4	5.0	5.2
C-by-C-MMSE	4.0	4.5	4.7
CRULE-MV	5.2	6.3	6.8
CRULE-AVE	8.0	8.3	8.6
ECRULE-MV	4.8	5.6	6.1
ECRULE-AVE	6.9	7.1	7.3
ECRULE-ICA	6.8	7.0	7.2

G, or B over that level were discarded. The three clipping levels used were 250, 300, and 400. We remind the reader that the illuminants are normalized so that perfect white has a maximum response of 255 among the three channels. The results with eight surfaces are shown in Table III. Additional results are available in [1].

Naturally, clipping degraded most algorithms, but the algorithms differ with respect to the degree of degradation. On the one extreme, clipping under these circumstances has little effect on the Color in Perspective algorithms. This is understandable because the level of clipping used was such that only specular pixels are clipped, and these pixels, having chromaticities near white, are not of much use to those algorithms.

At the other extreme, clipping essentially disables the ability of SCALE-BY-MAX to use specularities to improve the illumination chromaticity estimate. As the number of surfaces increases (see [1] for a plot), the effect becomes less damaging because it becomes more likely that there is at least one specular pixel just below the clipping level, which helps the algorithm. However, as the number of surfaces becomes very large, SCALE-BY-MAX degrades quite rapidly. This is in contrast to all other algorithms tested, where the negative effect of clipping decreases monotonically with the number of synthetic surfaces.

The reason for the increase in SCALE-BY-MAX error with a large number of surfaces and therefore many random specularities, is that the nondiscarded pixels now tend to have a maximum in each channel that approaches the clipping level. For example, in the case of the clipping level of 300, then the illuminant estimate will converge to (300, 300, 300). This is the same answer as the NOTHING algorithm! It is important to note that even though we do not normally have this diversity of colors in real images, a similar effect does in fact occur in images with fewer colors but with a wide variety of strengths of specularities.

VII. CONCLUSIONS

All algorithms studied here make some assumptions about the statistics of the reflectances to be encountered, and most make assumptions about the illuminants that will be encountered. The gray world algorithms make assumptions about the stability of the expected value of scene averages; SCALE-BY-MAX makes a similar assumption about the maximum in each channel; the gamut mapping algorithms

make assumptions about the ranges of expected reflectances and (for some variants) illuminants. Each method for choosing the solution makes additional assumptions. The neural net method and color by correlation methods go further and model the occurrence distributions. As assumptions get stronger, the prospect for success increases. It remains an open question to the extent that vulnerability to failures of the assumptions also increases. We seek algorithms which can exploit reasonable assumptions, preferably backed by empirical studies, but which are not overly sensitive to common failures of these assumptions.

Our experiments indicate that the methods which emphasize the use of input data statistics, specifically Color by Correlation and the neural net algorithm, are potentially the most effective at estimating the chromaticity of the scene illuminant. Some of the 3-D variants of Forsyth's gamut-mapping method also do well, and these algorithms have the advantage that they are able to also estimate the illuminant brightness. When specularities are present, these methods do even better, emphasizing that even when only chromaticity is of interest, full color algorithms should be considered.

Our detailed study of the effect of specularities on algorithms showed that their effect is significantly algorithm dependent. We also found that the effect of subsequent clipping of specular values is again algorithm dependent. Specularities are very common and are often clipped in standard cameras, especially when the aperture is automatically controlled. On the other hand, high dynamic range systems for robotics and high quality imaging will want to take advantage of the illuminant information carried by specularities. Thus, our study is relevant to both applications.

The work in this paper has laid the foundation for future work with image data. We have developed a comprehensive understanding on how a number of the leading algorithms perform in controlled circumstances, and we are therefore in an excellent position to interpret results from image data obtained with complementary methodology. Results of such a study will be made available in part two of this paper.

REFERENCES

[1] K. Barnard, "Practical color constancy," Ph.D. dissertation, Simon Fraser Univ. Sch. Comput. Sci., [Online]. Available: ftp://fas.sfu.ca/pub/cs/theses/1999/KobusBarnardPhD.ps.gz, Burnaby, BC, Canada, 1999.

[2] Data for computer vision and computational color vision. [Online]. Available: http://www.cs.sfu.ca/~color/data

[3] Implementations for selected color constancy algorithms. [Online]. Available: http://www.cs.sfu.ca/~color/code

[4] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Res. Applicat.*, vol. 27, no. 3, pp. 147–151, 2002.

[5] L. T. Maloney and B. A. Wandell, "Color constancy: A method for recovering surface spectral reflectance," *J. Opt. Soc. Amer. A*, vol. 3, pp. 29–33, 1986.

[6] G. Buchsbaum, "A spatial processor model for object color perception," *J. Franklin Inst.*, vol. 310, pp. 1–26, 1980.

[7] R. Gershon, A. D. Jepson, and J. K. Tsotsos, "From $[R, G, B]$ to surface reflectance: Computing color constant descriptors in images," *Perception*, pp. 755–758, 1988.

[8] D. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, pp. 5–36, 1990.

[9] J. A. Worthey, "Limitations of color constancy," *J. Opt. Soc. Amer. [Suppl.]*, vol. 2, pp. 1014–1026, 1985.

[10] J. A. Worthey and M. H. Brill, "Heuristic analysis of von Kries color constancy," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1708–1712, 1986.

[11] G. West and M. H. Brill, "Necessary and sufficient conditions for von Kries chromatic adaptation to give color constancy," *J. Math. Biol.*, vol. 15, pp. 249–258, 1982.

[12] G. D. Finlayson, "Coefficient color constancy," Ph.D. dissertation, Simon Fraser Univ., School of Computing, Burnaby, BC, Canada, 1995.

[13] G. D. Finlayson, M. S. Drew, and B. V. Funt, "Spectral sharpening: Sensor transformations for improved color constancy," *J. Opt. Soc. Amer. A*, vol. 11, pp. 1553–1563, 1994.

[14] K. Barnard, F. Ciurea, and B. Funt, "Sensor sharpening for computational color constancy," *J. Opt. Soc. Amer. A*, vol. 18, pp. 2728–2743, 2001.

[15] G. D. Finlayson, "Color in perspective," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 1034–1038, Oct. 1996.

[16] G. D. Finlayson, P. H. Hubel, and S. Hordley, "Color by Correlation," in *Proc. IS&T/SID 5th Color Imaging Conf.: Color Science, Systems, and Applications*, 1997, pp. 6–11.

[17] G. Finlayson and S. Hordley, "A theory of selection for gamut mapping color constancy," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp. 60–65.

[18] B. Funt, V. Cardei, and K. Barnard, "Learning color constancy," in *Proc. IS&T/SID 4th Color Imaging Conf.: Color Science, Systems and Applications*, Scottsdale, AZ, 1996, pp. 58–60.

[19] S. A. Shafer, "Using color to separate reflection components," *Color Res. Applicat.*, vol. 10, pp. 210–218, 1985.

[20] M. D'Zmura and G. Iverson, "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces," *J. Opt. Soc. Amer. A*, vol. 10, pp. 2148–2165, 1993.

[21] H.-C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1964–1699, 1986.

[22] S. Tominaga and B. A. Wandell, "Standard surface-reflectance model and illuminant estimation," *J. Opt. Soc. Amer. A*, vol. 6, pp. 576–584, 1989.

[23] W. M. Richard, "Automated detection of effective scene illuminant chromaticity from specular highlights in digital images," M.Sc. thesis, Center for Imaging Science, Rochester Inst. Technol., Rochester, NY, 1995.

[24] S. Tominaga, "Realization of color constancy using the dichromatic reflection model," in *Proc. IS&T/SID 2nd Color Imaging Conf.: Color Science, Systems, and Applications*, 1994, pp. 37–40.

[25] M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Res. Applicat.*, vol. 19, pp. 4–9, 1994.

[26] E. L. Krinov, *Spectral Reflectance Properties of Natural Formations*. Ottawa, ON, Canada: Nat. Res. Council of Canada, 1947.

[27] K. Barnard and B. Funt, "Camera characterization for color research," *Color Res. Applicat.*, no. 3, pp. 152–163, 2002.

[28] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *J. Opt. Soc. Amer. A*, vol. 14, pp. 1393–1411, 1997.

[29] G. D. Finlayson, B. V. Funt, and K. Barnard, "Color constancy under varying illumination," in *Proc. 5th Int. Conf. Computer Vision*, 1995, pp. 720–725.

[30] G. Finlayson, S. Hordley, and P. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 1209–1221, Nov. 2001.

[31] P. M. Hubel and G. Finlayson, "White point estimation using correlation matrix memory," U.S. Patent 6 038 339, 2000.

[32] K. Barnard, L. Martin, and B. Funt, "Color by correlation in a three dimensional color space," in *Proc. 6th Eur. Conf. Computer Vision*, Dublin, Ireland, 2000, pp. 375–389.

[33] G. Sapiro, "Color and illuminant voting," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 1210–1215, Nov. 1999.

[34] J. J. McCann, S. P. McKee, and T. H. Taylor, "Quantitative studies in Retinex theory," *Vis. Res.*, vol. 16, pp. 445–458, 1976.

[35] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, pp. 108–129, 1977.

[36] J. J. McCann, "Magnitude of color shifts from average quanta catch adaptation," in *Proc. IS&T/SID 5th Color Imaging Conf.: Color Science, Systems and Applications*, 1997, pp. 215–220.

[37] E. H. Land, "Recent advances in Retinex theory and some implications for cortical computations: Color vision and the natural image," *Proc. Nat. Acad. Sci.*, vol. 80, pp. 5163–5169, 1983.

[38] ——, "Recent advances in Retinex theory," *Vis. Res.*, vol. 26, pp. 7–21, 1986.

[39] D. A. Brainard and B. A. Wandell, "Analysis of the Retinex theory of color vision," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1651–1661, 1986.

[40] E. H. Land, "An alternative technique for the computation of the designator in the Retinex theory of color vision," *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3078–3080, 1986.

[41] B. K. P. Horn, "Determining lightness from an image," *Comput. Vis., Graph., Image Process.*, vol. 3, pp. 277–299, 1974.