A Comparison of Computational Color Constancy Algorithms—Part II: Experiments With Image Data

Kobus Barnard, Lindsay Martin, Adam Coath, and Brian Funt

***Abstract*—We test a number of the leading computational color constancy algorithms using a comprehensive set of images. These were of 33 different scenes under 11 different sources representative of common illumination conditions. The algorithms studied include two gray world methods, a version of the Retinex method, several variants of Forsyth's gamut-mapping method, Cardei *et al.*'s neural net method, and Finlayson *et al.*'s Color by Correlation method. We discuss a number of issues in applying color constancy ideas to image data, and study in depth the effect of different preprocessing strategies. We compare the performance of the algorithms on image data with their performance on synthesized data. All data used for this study is available online 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).**

Experiments with synthesized data (part one of this paper) suggested that the methods which emphasize the use of the 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. Unfortunately, we were unable to realize comparable performance on real images. Here exploiting pixel intensity proved to be more beneficial than exploiting the details of image chromaticity statistics, and the three-dimensional (3-D) gamut-mapping algorithms gave the best performance.

***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. This leads to a problem for many applications where the main interest is in the physical content of the scene. Consider, for example, a computer vision application which identifies objects by color. If the colors of the objects in a database are specified for tungsten illumination (reddish), then object recognition can fail when the system is used under the very blue illumination of blue sky. This is because the change in the illumination affects object colors far beyond the tolerance required for reasonable object recognition. Thus the illumination must be controlled, determined, or otherwise taken into account.

The wide applicability of separating a captured signal into parts which are due to the world, and parts which are due to the illumination, has lead to much interest in computational methods for doing so. In this paper we test a variety of promising approaches to this problem, using a large set of carefully calibrated images (all data is available online [1], as are implementations for most of the algorithms [2]). Applying the algorithms to images of actual scenes taken with a physical camera leads to the issues of camera characterization, image preprocessing, and the suitability of illuminant and reflectance sets used for algorithm calibration (training) to the images that will be encountered by the camera. Currently there is no satisfactory characterization of the images that an arbitrary vision system will encounter. Therefore an important part of this work is to explore the effects of reasonable mismatches between the assumptions made for algorithm calibration and reality. Our main vehicle for this is a comparison of results on image data with those on synthesized data from a companion paper [3].

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).

L. Martin, A. Coath, and B. Funt are with Simon Fraser University, Burnaby, BC, Canada. (e-mail: funt@cs.sfu.ca).

Publisher Item Identifier 10.1109/TIP.2002.802529.

II. Assumptions and Context

We do not test any algorithms which specifically require specularities to be present. However, as shown in part one, algorithms developed in the context of matte reflection vary substantially in their performance change due to specularities, and thus images with specularities is of interest. Furthermore it is difficult to take a comprehensive set of images without some specular reflectances. Therefore the images in our data set have varying amounts of dielectric specularities.

We assume that the illumination is constant across the scene. In synthetic experiments this is easy to enforce, but with image data it is a significant potential confound. Even when a single light source is used, this assumption is only grossly correct in the case of the illumination intensity. Sources of spatial variance in illumination intensity include geometric effects such as shading, illumination effects such as spatially extended light sources, and optical effects such as vignetting [4, p. 26] and a fall-off proportional to the fourth power of the cosine of the off axis angle [4, p. 208]. The optical effects could be calibrated for with sufficient effort, but such a calibration would be a function of camera settings such as focus and zoom, and doing so would not solve the bulk of the problems; therefore these issues have been ignored for this study.

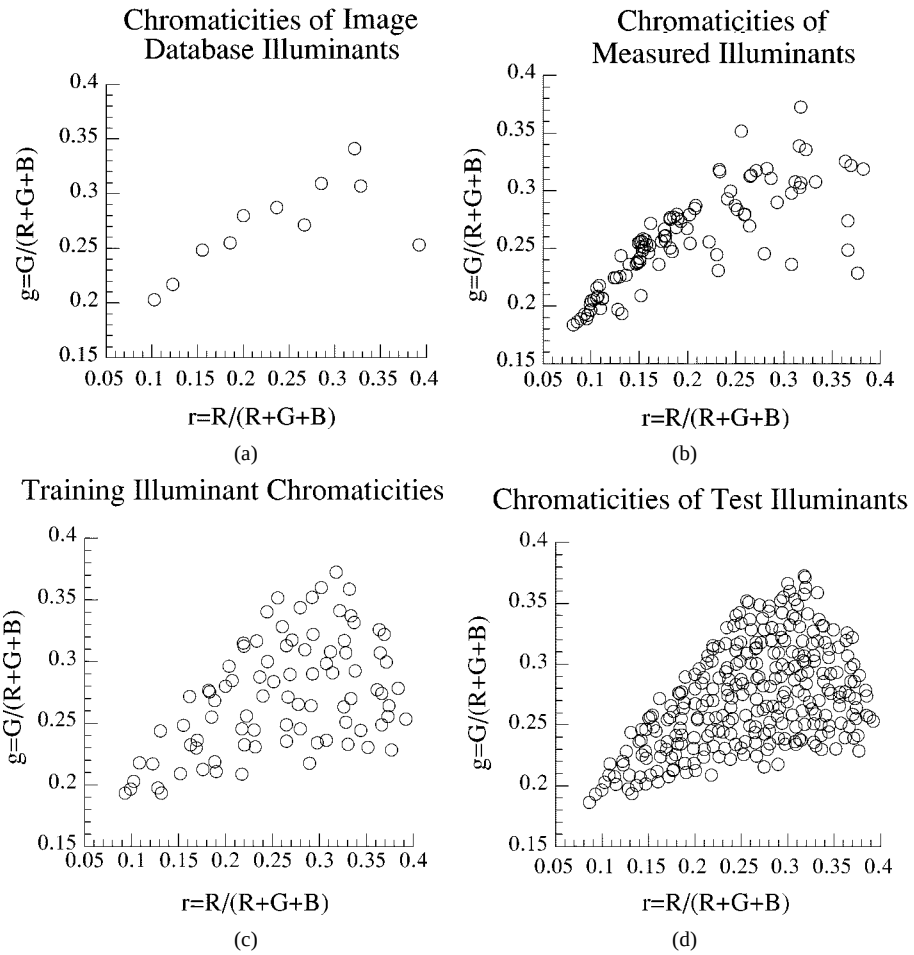

Fig. 1. Chromaticity distributions of the illuminant sets. The 11 illuminants used for creating test images are shown in (a). In (b), we plot the chromaticities of a set composed of additional sources and a number illuminations measured in and around our university campus. The training set constructed from these sources is shown in (c). For comparison (d) shows the illuminant set used for testing with synthesized data.

In the case of illuminant chromaticity, the assumption of spatial uniformity is generally more valid. However, inter-reflection can still cause major deviations. Furthermore, in natural images there are often multiple sources of illumination. For example, in outdoor images the sun and the sky often illuminate different parts of a scene with varying strengths. Color constancy under such conditions is beyond the scope of this paper (the interested reader is referred to [5]–[10]).

Color constancy algorithms also generally make assumptions about the diversity, and sometimes the detailed statistics, of the surfaces and the illuminants that may be encountered. Typically the surfaces and illuminants are supplied as collections of surface reflectances and illuminant energy spectra, and the assumptions an algorithm makes about them are manifest in calibration (training) data sets. For surface reflectances we used a set of 1995 spectra compiled from several sources (see part one for details [3]). This set was chosen to be a superset of the reflectance sets used by others for color constancy research. The range of color largely encompasses that found in our image database.

The illuminant spectra for algorithm calibration were chosen to roughly uniformly cover the (r, g) chromaticities of common illumination conditions. All illuminant spectra were normalized so that our cameras response to perfect white would have a maximum response among the three channels of 255. The 11 sources used to capture the image data were also selected to roughly cover the range of common illuminant conditions (see [3] for details). The chromaticities of the illuminant sets are shown in Fig. 1. Where relevant we use the Sylvania 50MR16Q (one of the 11 sources) as the standard (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.

III. Camera Characterization

We characterized our Sony DXC-930 CCD camera as described in [11], and used these sensors for generating camera responses for algorithm calibration. In experiments with synthesized data, where the same camera model is used both for training/calibration and data generation, the degree to which the model properly characterizes an actual camera is not critical. By contrast, when the algorithms are applied to data taken with a physical camera, characterization issues are important. It is possible to construct color constancy algorithms using only measured data, and doing so avoids characterization errors. However, it is more efficient to obtain camera independent quantities (such as reflectance functions and illuminant spectra) and update the camera model, rather than collect data for each camera (but see [12] for a compromise).

We model the ideal (linearized) camera response for channel k, $\rho^{(k)}$, for a surface with reflectance spectra $S(\lambda)$ under an illuminant with spectra $E(\lambda)$ by

$$\rho^{(k)} = F^{(k)}\left(v^{(k)}\right) = \int L(\lambda)R^{(k)}(\lambda)\,d\lambda \qquad (1)$$

where $R^{(k)}$ is a sensor sensitivity function for the kth channel, $F^{(k)}$ is a wavelength independent linearization function, and $\rho^{(k)}$ is a the linearized camera response. In practice, the functions of wavelength 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.

Most color constancy algorithms assume that the image pixels are proportional to the input spectral power which is equivalent to assuming either that $F^{(k)}$ is the identity function, or that is known and has been applied. Since $F^{(k)}$ is very often not the identity, when we apply color constancy to data from a real camera we must determine it as part of the camera characterization step. This can be done as an initial phase of characterization [13], [14] or as an integral part where $R^{(k)}$ and $F^{(k)}$ are determined jointly [11]. The camera sensors (Fig. 2), as well as the data used to estimate them, are available online [1].

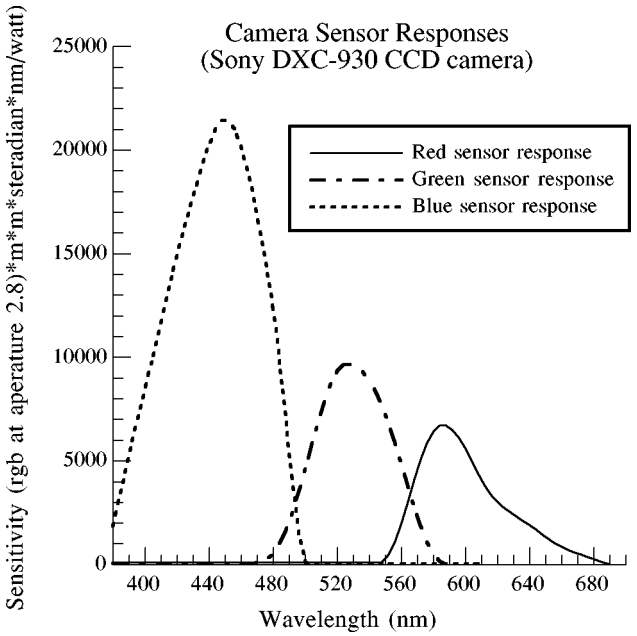

Fig. 2. Camera sensor response functions.

IV. Diagonal Color Constancy

We assume the diagonal model of illumination change which maps the image taken under one illuminant, to the image taken under another illuminant, by simply 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 $\rho^U = (\rho_1^U, \rho_2^U, \rho_3^U)$, and that the response under a known, canonical, illuminant is $\rho^C = (\rho_1^C, \rho_2^C, \rho_3^C)$. Then the response of the white patch can be mapped from the unknown case to the canonical case simply by scaling the ith channel by ρ_i^C/ρ_i^U. 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] [15]–[18]. 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, performance may be improved by using sensor sharpening [19], [20].

V. Algorithms

Table I summarizes the algorithms chosen for study. A more comprehensive introduction is provided in part one of this paper [3]. Here, we briefly outline the algorithms and provide some details specific to their use in this study. Implementations for most of these algorithms are available online [2].

A. Gray World and Illumination Estimation by the Maximum of Each Channel

The gray world method assumes that the average of the observed image (R, G, B) is a good estimate of the camera response to "gray." Ideally "gray" is defined by the expected average over the application domain. This is not generally available, and in this work we consider two algorithms based on alternatives. The first (GW) uses a 50% uniform reflectance for gray, and the second (DB-GW) uses the average of the reflectance spectra in the reflectance dataset defined above. For GW the average scene (R, G, B) is converted to an illuminant estimate by scaling it by a factor of two. For DB-GW we apply the diagonal model and 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.

The limiting case of one Retinex algorithm, SCALE-BY-MAX, simply estimates the illuminant (R, G, B) by the maximum response in each channel [7], [21]–[23].

B. Gamut Mapping Methods

We present the results of a number of algorithms based on Forsyth's gamut-mapping approach [18], [24]–[28]. Here we consider the set of all possible (R, G, B) due to surfaces in the world under the known, "canonical" illuminant. This set is convex and is represented by its convex hull. The set of all possible (R, G, B) under the unknown illuminant is similarly represented by its convex hull. Under the diagonal assumption of illumination change, these two hulls are a unique diagonal mapping (a simple three-dimensional (3-D) stretch) of each other. Because the observed set is normally a proper subset, the mapping to the canonical is not unique, and Forsyth [24] provides a method for effectively computing the set of possible diagonal maps which is convex set in the space of mapping coefficients (see [3], [24], [28] for details). Finlayson's Color in Perspective algorithm adds two additional ideas [26]. First, the gamut-mapping method can be used in the chromaticity space $(R/B, G/B)$. Second, the diagonal maps can be further constrained by restricting them to ones corresponding to expected illuminants.

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

TABLE I
KEY TO THE ALGORITHM LABELS USED IN THIS PAPER. ALL ALGORITHMS ARE DISCUSSED IN DETAIL IN [3]. ADDITIONAL REFERENCES ARE PROVIDED IN COLUMN THREE

Label	Description	References
NOTHING	Output image is input image (illuminant is canonical illuminant)	
AVE-ILLUM	Illuminant RGB is set to an average over common illuminant RGB's	
GW	Gray World: Illuminant RGB is twice average of image RGB	
DB-GW	Data Base Gray World: Illuminant RGB is ratio of average image RGB to RGB of average of the reflectance database.	
SCALE-BY-MAX	Illuminant R is the maximum of image R, and similarly for G and B	
CIP-MV	Color in Perspective (gamut mapping in 2D) using the maximum volume heuristic to choose the solution	[26]
CIP-AVE	Color in Perspective using average of solution set	[25]
CIP-ICA	Color in Perspective with solution averaging in three dimensions	[27]
NEURAL-NET	Neural net trained to estimate illuminant (r,g)	[33, 34]
C-by-C-01	Color by Correlation without Bayesian statistics	[29]
C-by-C-MAP	Color by Correlation with MAP (maximum a posteriori estimate)	[30]
C-by-C-MLM	Color by Correlation with MLM (maximum local mass estimate)	[30, 32]
C-by-C-MMSE	Color by Correlation with MAP (minimum mean square error estimate)	[30]
CRULE-MV	Gamut mapping using maximum volume heuristic to choose solution	[24]
CRULE-AVE	Gamut mapping using average of diagonal maps to choose solution	[25]
ECRULE-MV	Gamut mapping in 3D with illumination constraint and maximum volume heuristic	[25]
ECRULE-AVE	Gamut mapping in 3D with illumination constraint and averaging the convex hull of the solution set to choose the solution	[25]
ECRULE-ICA	Gamut mapping in 3D with illumination constraint and using the center of mass of the non-convex solution set to choose the solution	[28]

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 with the illumination constraint, designated by CIP, and the illumination constraint set applied to CRULE designated by ECRULE (for "extended-CRULE"). These solution sets are paired with three methods of selecting a solution from them. MV denotes the original maximum volume heuristic which is simply the diagonal transform with maximal determinant; AVE specifies that the constraint set is averaged, using a convex approximation to the illumination constraint if necessary; and ICA specifies that the constraint set is numerically integrated to deal with the fact that it is non-convex ("illumination constrained average").

The canonical gamut, the canonical illuminant (R, G, B), and the illuminant set are all derived from the calibration sets mentioned above, and described further in part one [3].

C. Color by Correlation

Recently, Finlayson *et al.* introduced Color by Correlation [29]–[31] as an improvement on the Color in Perspective method. The basic idea of Color by Correlation is to precompute 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 [29]. The first version of this algorithm (C-by-C-01) is essentially an alternative implementation of Color in Perspective.

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 by one various methods. Here we consider the maximum likelihood (C-by-C-MAP), mean likelihood (C-by-C-MMSE), and the local area mean (C-by-C-MAP), introduced in [32]. 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. We report results for a sigma of 8 (see part one [3] for further discussion).

D. Neural Net Methods

The results labeled NEURAL-NET are from a neural network trained to estimate the color of the illuminant [33]–[35]. The neural net is a multilayer Perceptron with two hidden layers. As is common, the general structure is pyramidal. 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. The output signal from the two output neurons are an estimate of the (r, g) chromaticity of the scene illuminant. 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. Extensive details are provided in [34] and [35].

VI. IMAGE DATA SET

We took pictures of 30 scenes under the 11 sources mentioned above, for a total of 330 images. Some of the images had to be culled due to problems, leaving 321 for our experiments. The scenes under one illuminant are shown in Fig. 3, and in Fig. 4 we plot some image statistics and the analogous statistics for

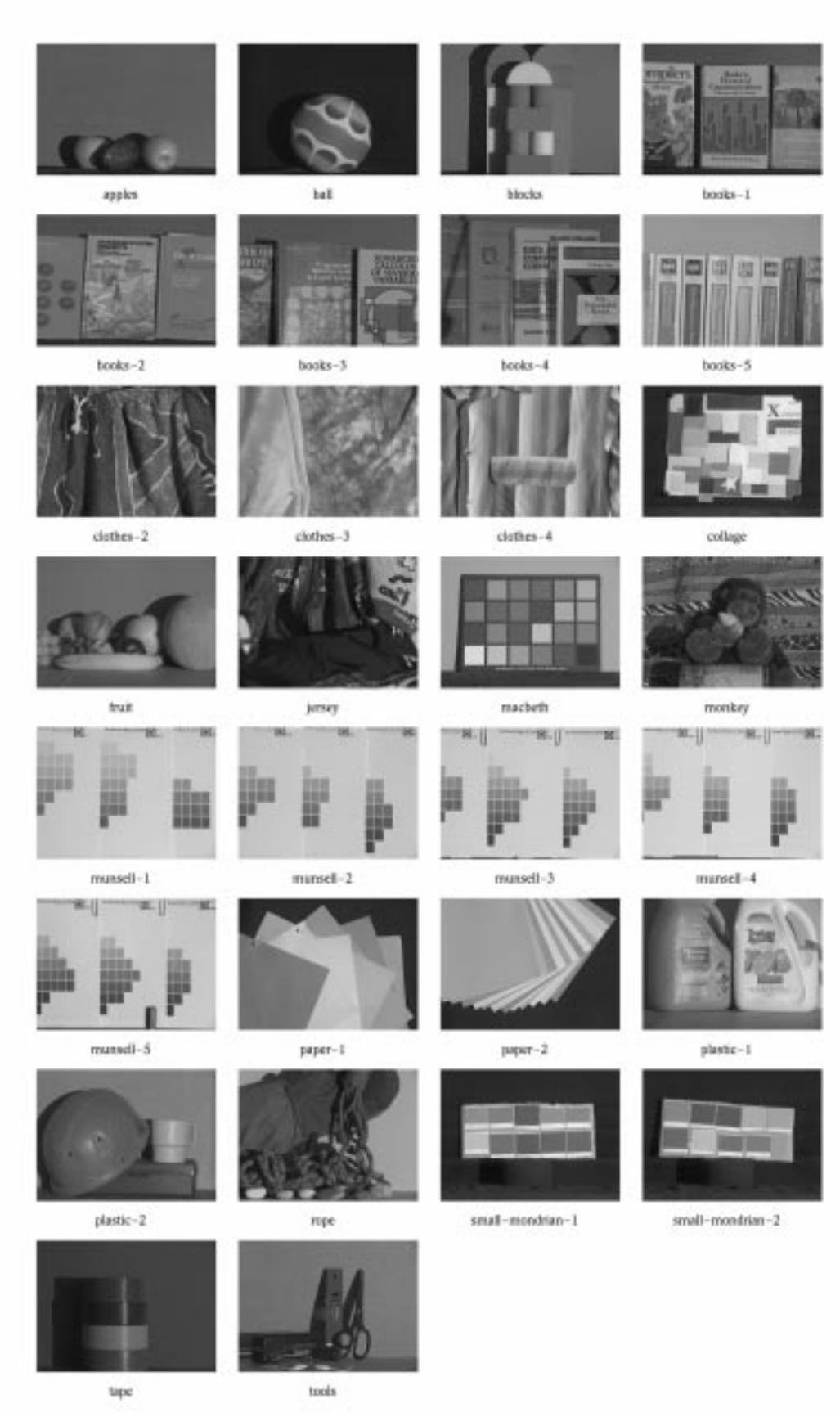

Fig. 3. The 30 scenes used in this study. Each scene was imaged under 11 illuminants.

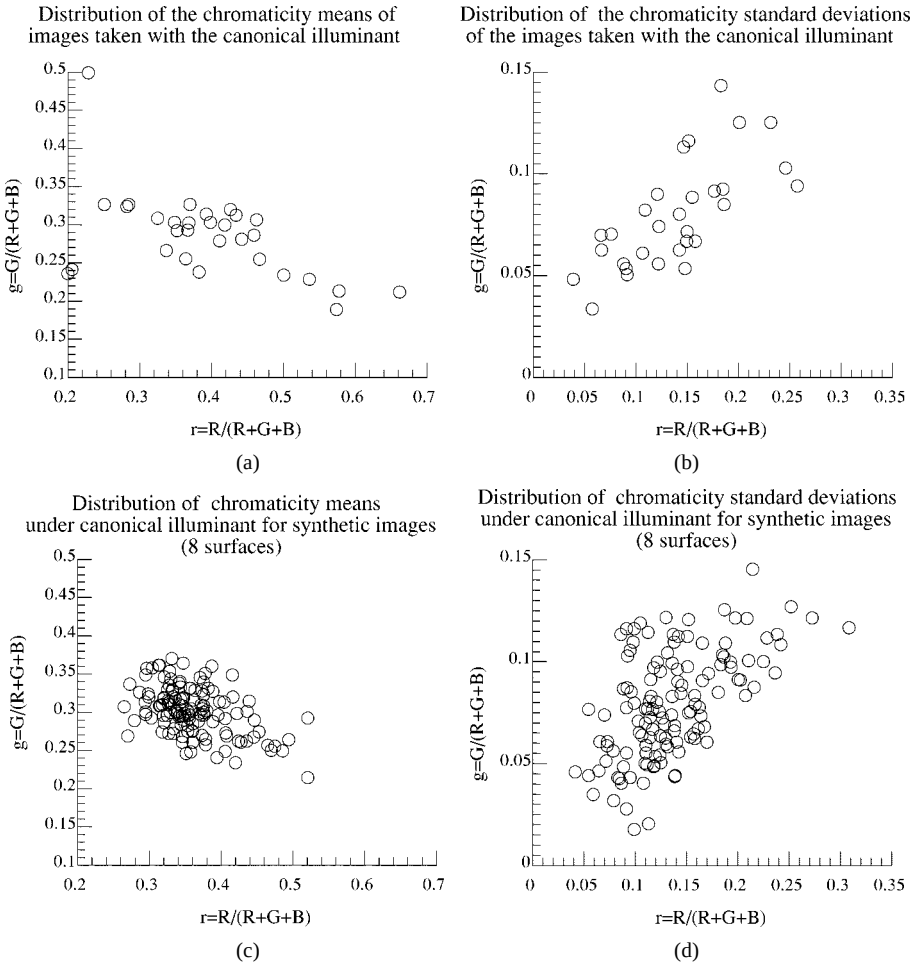

Fig. 4. Distribution of the (a) means and (b) standard deviations of the chromaticities of scenes as imaged under the canonical illuminant and preprocessed with the generic method used for the results in Table I. The means and standard deviations for the 1000 synthesized scenes with eight surfaces used for comparison are shown in (c) and (d).

synthetic data. The images had varying amounts of specularities, but we avoided images with colored metals or fluorescent surfaces.

The experimental routine was as follows. First, a new scene was constructed. We then placed a reference white standard in the center of the scene, perpendicular to the direction of the illuminant. The position of the illuminant was set so that the number of clipped pixels was small. This meant that if the scene had bright specularities, then the image was purposely underexposed. We then took a picture of the scene with the reference white in the center. Finally, we removed the reference white, and took 50 successive pictures which were averaged to obtain the final input image. We then repeated the process for the remaining ten illuminants, and then we moved onto the next scene.

The images with the reference white were used to provide the "answer." We extracted the central 30 by 30 pixel window of each of these images, and used the average (R, G, B) over these windows as the estimate of the illuminant for the corresponding input images. Both the input images and the illuminant (R, G, B) estimate were mapped into a more linear space as determined by the camera characterization process [11]. In addition, a spatially varying color cast due to the camera optics was removed, as well as some fixed pattern noise. Additional details of the image capture and subsequent corrections can be found in [28] in [36].

After the processing described above, the images were nearly linear with scene radiance, and significantly cleaner than single shot images taken with the same camera. Averaging reduces independent Gaussian noise by the square root of the number of samples, so the expected improvement here is a factor of about seven. We note that averaging does not help with some sources of noise such as lens flare. We found that the images could be scaled by a factor of ten without incurring too much noise—visually comparable to a normally exposed image—which is consistent with the expected improvement. Thus the images could be scaled and clipped to emulate capture with an automatic aperture, as well as capture with a higher dynamic range device. The bit depth of the images is roughly 9–10, as compared to 6–7 for our single frame images where the noise level is of the order of 2.5. After linearization and other adjustments, the range of the image gray levels was roughly 0 to 240.

The (R, G, B) of the white standard provided a good estimate of the chromaticity of the illuminant, but the error in the illuminant magnitude for any given picture could be quite high—easily 10%, because of the difficulties in keeping the white reflectance standard perpendicular to the light source. Furthermore, three of the sources were distended, and here we simply attempted to find the orientation which maximized the brightness of the reflectance standard.

Some algorithms are sensitive to small values in one or more channels, and therefore we excluded any pixel whose R, G, or B, was 2 or less, after the linearization and other corrections discussed above had been applied. We also excluded any pixel with R, G, or B over 240 after linearization (corresponds to roughly 250 before linearization), as these pixels may be clipped. We used a slightly lower value than the actual clipping level because the frame averaging process and/or noise may hide clipping. Since we avoided clipping as much as possible during the capture process, this exclusion applied to only a few pixels.

VII. Preprocessing for Computational Color Constancy

All the algorithms discussed above are developed in terms of one input (R, G, B) per identifiable image surface. This is a little different than a real image, which is a collection of multiple samples per surface, including samples which straddle surface boundaries. Many algorithms are indifferent to the statistics of the sampling; only the presence or absence of a color is relevant. Thus it is common to simply use the image pixels themselves as input. However, initial experiments indicated that it is better to first spatially average the images, with a block size of 5×5 being roughly optimal for our camera. Nonetheless, blindly averaging the image in this way reduces the information available to the algorithms, and such a step should be less important when frame averaging is used. Interestingly, we found that this was not case.

We believe that the reason for this anomaly is as follows: using each pixel as a datum makes the implicit assumption that the (R, G, B) of pixels straddling two surfaces is a convex combination of the (R, G, B)s of the pixels to either side. Careful examination of images reveals that, for our camera, this assumption does not hold, possibly due to chromatic aberration or misregistration of the CCD elements (our camera has three CCD arrays—the incoming signal goes through beam splitters and filters on its way to the CCD arrays). Specifically, we found that the (R, G, B)s of a nonnegligible number of boundary pixels were not the convex combination of surrounding pixels, and thus should be considered erroneous data. Furthermore, cameras with mosaic'ed sensors, which are more common than three CCD cameras, are also susceptible to similar anomalies.

This problem, together with the above observation that the algorithms are expressed in terms of surfaces, leads to the consideration of image segmentation as a form of preprocessing, and we investigated this idea in detail in our experiments. We were able to find segmentation parameters which improved most algorithms, although the effect was quite algorithm dependent, and in fact, more volatile than we expected. We present results using these, somewhat arbitrarily chosen, general purpose parameters. However, since it is reasonable for a proponent of a given algorithm to optimize the preprocessing for that algorithm, we also present results where the optimal among all preprocessing methods was chosen on an algorithm-by-algorithm basis.

To segment the images we used region-growing, subject to two constraints. First, we ensured that all chromaticities in a region were within a certain absolute tolerance of each other. Second, we ensured that the pixel brightness, quantified by $R+G+B$, of all pixels in a region were within a certain relative tolerance of each other. Thus adjacent pixels (horizontally or vertically connected) were added to regions, beginning with unassigned pixels as region seeds, provided that these constraints were met. In addition, we insisted that the region was larger than a certain number of pixels. Thus there were three segmentation parameters. We used four different values of the chromaticity tolerance (0.0025, 0.005, 0.01, 0.02), four for the relative brightness range (10%, 20%, 30%, and 40%), and three for the minimum number of pixels (5, 10, 20). The (R, G, B) averages of each region was then used as input to the algorithms. A second preprocessing strategy was to average the pixels in image blocks with lengths of sides (1, 2, 3, 5, and 7). A third strategy was to then put block averaged result into bins in (R, G, B) space (100 divisions per channel), and use the average of the (R, G, B) in each of the bins, thereby removing duplicate colors from the input. A final preprocessing approach was to use the (R, G, B) vertices of the convex hull of the data instead of the data itself. This is motivated by the knowledge that for the gamut-mapping algorithms, the hull boundary points are the ones that matter. Of course, using the convex hulls has no effect on gamut-mapping algorithms, but preliminary results with the neural net and Color by Correlation were promising. Therefore we also tested using this in conjunction with block averaging (five-pixel blocks) and two selected segmentation methods. Thus a total of 65 preprocessing methods were tested.

VIII. Experiments

In Table II we present the results using a generic preprocessing method that works relatively well with most algorithms, but is optimal for none of them. Specifically, the images were segmented subject to the constraint that the (r, g) vector distance of any two pixels in the region was not more than 0.005, that the value of $R + G + B$ did not vary by more than 10%, and that the region had at least five pixels. Once segmented, the average of (R, G, B) over each of the regions was used as the input to the algorithms. Table II also includes results using input modified to emulate our camera when the automatic aperture is active. The image data was artificially scaled and clipped so that the maximum (R, G, B) of the reference white would be 300, and all pixels with R, G, or B over 255 were discarded. In Table III we provide the results of our preprocessing experiments. Here we show the range of results obtained using the 65 preprocessing methods described above.

In Fig. 5, we compare the results on the image data, using the best preprocessing method for each algorithm, with results on synthesized scenes with eight surfaces (from [3]). Absolute errors found with generated and captured data are not generally comparable, but to study the changes in relative performance we identified eight synthetic surfaces as been roughly as difficult as our images. Specifically, in Fig. 5, the average error across the algorithms is roughly the same on images as on eight synthetic surfaces. We use "comparable difficulty" as the calibration point to reduce possible confounds due to "problem difficulty" in our comparisons across generated and captured data results.

TABLE II
ALGORITHM PERFORMANCE FOR 321 REAL IMAGES USING A GENERIC PRE-PROCESSING METHOD (SEE TEXT). UNCERTAINTY IS ROUGHLY 4%. THE TWO RIGHTMOST COLUMNS ARE THE CHROMATICITY RESULTS OBTAINED ON IMAGES SCALED AND CLIPPED TO EMULATE THE DATA FROM OUR CAMERA WHEN THE AUTOMATIC APERTURE IS USED

	Extended dynamic range				Standard dynamic range emulated by clipping the input data	
Algorithm	Illuminant estimate angular error	Illuminant estimate rg error	Illuminant estimate RGB error	Illuminant estimate R+G+B error	Illuminant estimate angular error	Illuminant estimate rg error
NOTHING	17.9	0.125	*	*	17.9	0.125
AVE-ILLUM	12.9	0.094	*	*	12.9	0.094
GW	13.8	0.109	154	248	13.9	0.109
DB-GW	11.7	0.094	115	175	11.8	0.094
SCALE-BY-MAX	8.9	0.060	93	143	9.1	0.062
CIP-MV	23.4	0.174	*	*	22.1	0.165
CIP-AVE	16.1	0.113	*	*	15.1	0.108
CIP-ICA	10.6	0.077	*	*	10.4	0.076
NEURAL-NET	9.5	0.070	*	*	9.5	0.070
SP-NEURAL-NET	9.1	0.068	*	*	9.1	0.068
C-by-C-01	10.9	0.081	*	*	10.5	0.078
C-by-C-MAP	9.9	0.072	*	*	9.9	0.072
C-by-C-MLM	9.9	0.072	*	*	9.9	0.072
C-by-C-MMSE	9.9	0.072	*	*	9.9	0.072
CRULE-MV	5.6	0.045	90	138	6.6	0.049
CRULE-AVE	7.1	0.053	278	428	8.3	0.063
ECRULE-MV	5.6	0.042	91	139	6.3	0.047
ECRULE-AVE	6.9	0.050	233	367	7.8	0.057
ECRULE-ICA	7.0	0.051	232	364	8.0	0.058

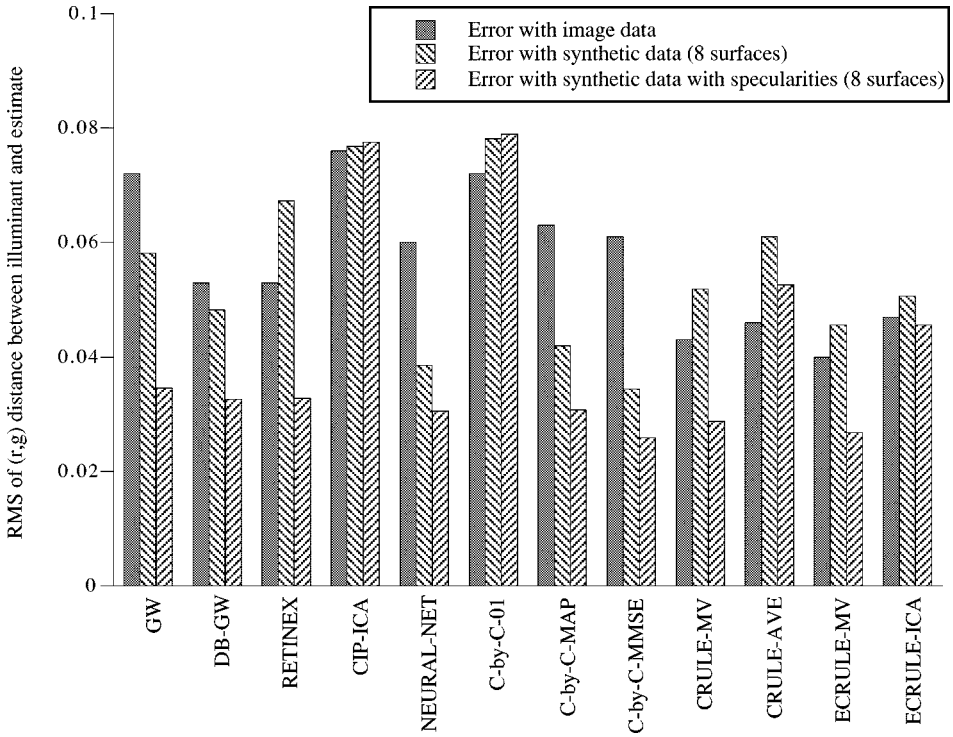

Fig. 5. Algorithm performance for 321 real images using the best preprocessing method for each algorithm compared with synthetic results for a comparably difficult number of surfaces (8).

The most significant deviation of the results with real images from those with synthesized data is that the algorithms which consider the detailed statistics of the input (Color by Correlation and the neural network) lose ground to the gamut-mapping algorithms in their ability to estimate the chromaticity of the illuminant. For example, on the image data, with optimal preprocessing, the (r, g) chromaticity error for C-by-C-MMSE was 80% that of CIP-ICA, down from 44% in the eight-surface synthetic case. Furthermore, the (R, G, B) gamut-mapping algorithms performed the best on the image data, compared

TABLE III
ALGORITHM CHROMATICITY PERFORMANCE FOR 321 REAL IMAGES USING THE BEST PRE-PROCESSING METHOD FOR THAT ALGORITHM. HENCE EACH RESULT IN THIS TABLE IS OBTAINED USING A (POTENTIALLY) DIFFERENT PRE-PROCESSING METHOD. THE SEGMENTATION PARAMETERS ARE (MIN-SEGMENT-SIZE, MAX-RG-VARIATION, MAX RELATIVE $R+G+B$ VARIATION). WE USE AVE(N) TO INDICATE THAT N BY N BLOCKS WERE AVERAGED

Algorithm	Minimum illuminant estimate rg error	Illuminant estimate rg error with generic method	Average illuminant estimate rg error	Maximum illuminant estimate rg error	Pre-processing method for minimum error
NOTHING	0.125	0.125	0.125	0.125	All are equal
AVE-ILLUM	0.094	0.094	0.094	0.094	All are equal
GW	0.072	0.109	0.103	0.137	Binarized RGB
DB-GW	0.053	0.094	0.086	0.123	Binarized RGB
SCALE-BY-MAX	0.053	0.060	0.063	0.079	Segmentation (5, 0.020, 0.3)
CIP-MV	0.149	0.174	0.185	0.200	Raw data
CIP-AVE	0.105	0.113	0.123	0.134	Raw data
CIP-ICA	0.076	0.077	0.080	0.085	Segmentation (5, 0.005, 0.3)
NEURAL-NET	0.060	0.070	0.068	0.081	Segmentation (20, 0.005, 0.4)
SP-NEURAL-NET	0.061	0.068	0.067	0.083	Segmentation (5, 0.020, 0.4)
C-by-C-01	0.072	0.081	0.078	0.088	Segmentation (20, 0.010, 0.4)
C-by-C-MAP	0.063	0.072	0.071	0.084	Segmentation (20, 0.005, 0.4)
C-by-C-MLM	0.062	0.072	0.070	0.083	Segmentation (20, 0.005, 0.4)
C-by-C-MMSE	0.061	0.072	0.070	0.082	Segmentation (20, 0.005, 0.4)
CRULE-MV	0.043	0.045	0.049	0.066	Segmentation (10, 0.005, 0.2)
CRULE-AVE	0.046	0.053	0.054	0.085	Segmentation (20, 0.010, 0.3)
ECRULE-MV	0.040	0.042	0.045	0.066	Segmentation (10, 0.005, 0.2)
ECRULE-AVE	0.046	0.050	0.051	0.079	Segmentation (20, 0.020, 0.4)
ECRULE-ICA	0.047	0.051	0.051	0.077	Segmentation (20, 0.020, 0.4)

with being somewhat worse than both the neural network and Color by Correlation methods in the eight-surface matte synthetic case. We will discuss this discrepancy in more detail below.

Our experiments with preprocessing show a significant effect on algorithm performance. The difference between the average method and the best method was usually greater than 10%, and in some cases, such as the two gray world algorithms, it was of the order of 30%. Furthermore, since this effect is quite algorithm dependent, a careful comparison of color constancy algorithms must take this into account. Our current strategy for doing this is to provide comparisons based on the optimal preprocessing chosen on an algorithm by algorithm basis as we do in Table III and Fig. 5.

Finally, we note that the simulated clipping we applied to this data did not have a major effect on the results. This is because overall, our image database did not include an over abundance of extreme specularities, and only about one third of the images had significant specularities. The effects that we did find are generally consistent with the results with synthesized data, except that SCALE-BY-MAX was degraded slightly less due to specularities compared to the 3-D gamut-mapping algorithms. However, overall, the impact of the clipping results on our conclusions is small, as it induces little change in the rank ordering of the algorithms. Thus our conclusions hold for moderately specular images, even if a significant number of those specularities are clipped. We remind the reader that this conclusion is based on excluding the clipped pixels, rather than optimistically using them, which is detrimental to several algorithms.

IX. DISCUSSION

The main discrepancy between our results on synthesized data and our results on image data is the drop in performance of Color by Correlation and the neural net method relative to the 3-D gamut-mapping methods. There are two principal differences which can explain performance variation between these two algorithm groups. First, the 3-D gamut-mapping methods can exploit pixel intensity information such as that inherent in specularities. Since the particular neural net and Color by Correlation approaches investigated here use chromaticity input, they cannot use this information. The second relevant difference is that the neural net and Color by Correlation can exploit the structure of the scene and illumination statistics; by contrast, gamut mapping uses simple statistics (average or maximum volume heuristic) to choose a solution from the constraint set. In our synthetic test domain, the advantage afforded by using the details of the statistics of the world more than made up for the disadvantage of not being able to use pixel intensity. On our image data, the reverse was true. In the following we will consider the extent to which this change is due to an increase in the relevance of pixel intensity, compared a decrease in the effectiveness of utilizing the detailed structure of image statistics.

We first consider specularities. When significant specularities were modeled in the synthetic data, 3-D gamut mapping performance approached that of the best algorithms in that domain (C-by-C-MMSE). Thus specularities can explain some of the difference. Not all difference can be explained because the specularities in the image dataset were not as extreme as the simulated ones. Furthermore, when we clipped the image data to emulate an automatic aperture, the 3-D gamut mapping algorithms still performed much better than C-by-C-MMSE.

This suggests that part of the difference is due to a mismatch in the statistics used for training compared to the statistics of our image dataset. Such a difference must be in the detailed structure of the statistics because gross features of the statis-

tics, such as the ranges of surfaces and reflectances, help both Color by Correlation and gamut mapping (Color by Correlation is a generalization of gamut mapping). Both algorithms choose solutions from the same constraint set. If the illuminant chromaticity is tightly constrained by gamut mapping, then Color by Correlation will also do well. Infeasible illuminants in the gamut mapping paradigm are zero probability solutions in the Color by Correlation framework. In fact, gamut mapping is roughly analogous to Color by Correlation with uniform statistics. For example, CIP-01 can be interpreted as an implementation of two-dimensional (2-D) gamut mapping. Since the 3-D algorithms do better, it is plausible that uniform statistics in (R, G, B) space are part of the reason.

To further explore this notion, we re-ran one of the real image experiments with correlation matrices computed using our reflectance data set augmented so that it yielded more uniformly distributed (R, G, B). We found that Color by Correlation did improve, but only by a small amount. Nonetheless, we observe that since C-by-C-MMSE did better than C-by-C-01 in all cases, both our reflectance data set, and the augmented one [uniform in (R, G, B)], are more appropriate for our image data set than the uniform (r, g) statistics implied by C-by-C-01.

To further clarify the nature of the discrepancy we undertook a second post-hoc experiment designed to provide the algorithms with the statistics of the image test set, and thereby remove the mismatch in training and testing statistics.[2] To do this, we calibrated both Color by Correlation and the gamut mapping algorithms using the image data. This strategy also reduces possible confounds due to inaccuracies in camera characterization. The data was extracted using the generic preprocessing method (as used for the results in Table II). The illuminant set was now limited to exactly the 11 sources, and the calibration data was no longer generated. Instead, the canonical gamuts were constructed from the (R, G, B) of all images under the canonical illuminant, and each row of the correlation matrices was constructed from the (R, G, B) of all the images under the corresponding illuminant. The results are shown in Table IV along with the corresponding spectral trained results from Table II.

As expected, calibration on image data decreased the error for all relevant algorithms. What was not expected was that Color by Correlation did not gain relative to gamut mapping. The error for the Color by Correlation algorithms was roughly 77% of that using spectral data compared with 70–83% for the 3-D gamut mapping variants. Color by Correlation performance did not even match that of gamut mapping calibrated with spectral data. Thus for this image data set, the exploiting of pixel brightness information is clearly more fruitful than exploiting image chromaticity statistics. It is key that in this final experiment there is no mismatch between the training and testing statistics. We also note that there is in fact some benefit to modeling the statistics of this data as demonstrated by the lower error of C-by-C-MMSE as compared to CIP-ICA. Thus the image data statistics are more favorable than uniform for Color by Correlation.

We are left with the significant open question of what is an appropriate characterization of the statistics for a practical application such as outdoor photography, and the extent to which the structure of those statistics can be exploited. The work here suggests that if the vehicle for doing so is a chromaticity based algorithm, the gain due to exploiting the detailed structure of image statistics will have to be relatively large to offset the advantage afforded by using the pixel intensities. A better strategy may be integrate the use of the detailed structure of image statistics into an algorithm which does not throw out intensity information. Several strategies for doing so have been recently proposed [37]–[39].

TABLE IV
CHROMATICITY PERFORMANCE OF SELECTED ALGORITHMS ON IMAGE DATA WHEN CALIBRATED USING SPECTRAL DATA (MIDDLE COLUMN) AND IMAGE DATA (RIGHTMOST COLUMN). THE RESULTS ARE FOR 321 REAL IMAGES USING THE GENERIC PREPROCESSING METHOD

Algorithm	Illuminant estimate angular error (degrees) when calibrated on spectral data	Illuminant estimate angular error (degrees) when calibrated on image data
CIP-ICA	10.6	9.8
C-by-C-MAP	9.9	7.6
C-by-C-MLM	9.9	7.6
C-by-C-MMSE	9.9	7.6
CRULE-MV	5.6	4.7
CRULE-AVE	7.1	5.5
ECRULE-MV	5.6	4.4
ECRULE-AVE	6.9	4.9
ECRULE-ICA	7.0	4.9

The use of pixel brightness information by the 3-D gamut-mapping algorithms bears further comment. In [27], Finlayson and Hordley show that, under reasonable assumptions,[3] the perspective gamut-mapping algorithms are as powerful as the 3-D ones in their ability to constrain illuminant chromaticity. In other words, all illuminants implied by the Color in Perspective constraint set are also present in the ECRULE constraint sets. Therefore, differences between the gamut-mapping algorithms are due to other considerations. For real data we cannot ignore robustness with respect to noise, but even with synthetic data we found a significant difference between the perspective and the 3-D gamut-mapping algorithms. Thus, the main difference between gamut-mapping algorithms is their ability to estimate the solution from the constraint set, and this is where the 3-D ones seem to be better suited.

In [3] we discuss how the presence of specularities can help 3-D gamut-mapping (and other) algorithms, and provided results with synthesized specularities with specific properties. A more general view of specularities is that they simply change the statistics of the input, and this change is most significant when pixel intensity is included. A second observation is that the 3-D gamut mapping algorithms also do well on synthetic data when there are no specularities, especially in comparison to their 2-D counter-parts. Therefore, we see specularities as especially good input for the 3-D gamut mapping algorithms, and their ability to exploit the range of full (R, G, B) information as the key to their good performance on our image data set.

[2]We thank one of the anonymous reviewers for suggesting this experiment.

[3]The assumption is that in the 3-D case, the origin is included in the canonical gamut. We agree that this is a reasonable assumption because surfaces may be arbitrarily dark due to shading.

We first analyze this ability to exploit pixel intensity beginning with solution selection by averaging. In 3-D constraint space, illuminants within a specific chromaticity range correspond to cones. Now consider an illuminant chromaticity in a small range near the edge of the chromaticity constraint set (see Fig. 6). In the perspective case, this small range is similar to any other of the same size. In the 3-D case, however, the corresponding cone has less volume than ones closer to the middle of the constraint set due to the typical shape of the constraint set. In general, the volume is a function of the shape of the 3-D constraint set, and solution selection by the two methods are not equivalent. Thus, to the extent that the shape encodes useful information, the perspective method is at a disadvantage.

With the maximum volume heuristic the difference between the two algorithms is even more extreme. There does not seem to be a workable analogy to this heuristic for perspective space (the naive one leads to biased algorithms). Thus again, the full color version of the gamut mapping algorithms can easily have an advantage over the perspective version for choosing the solution, even when the final goal is only chromaticity estimation. This is supported empirically, as well as theoretically in the case of bright specularities.

In summary, the performance of the algorithms on our image data, relative to that on our synthesized data, was due to a combination of the proposed factors. Some of the difference was explained by the mismatch of the calibration statistics and the image statistics; more difference was due to the statistics being less exploitable by Color by Correlation in the case of image data; and the most difference was explained by the pixel intensities being more exploitable by the 3-D gamut mapping algorithms in the image data case.

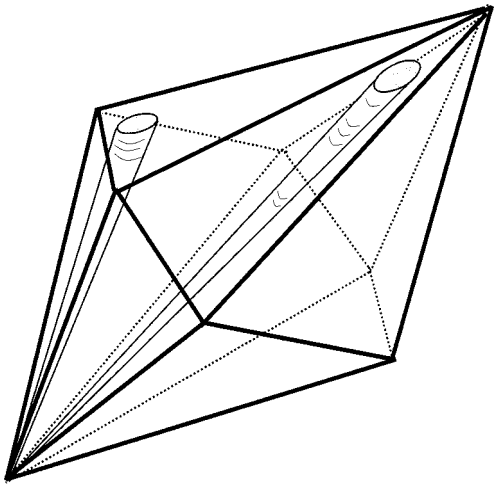

Fig. 6. Illustration of the basic shape of a simple constraint set in mapping space due to observed (R, G, B) for the 3-D gamut mapping methods. The illumination constraint is not shown. The figure shows that there is structure beyond that accessible by the cones implied by the perspective simplification. For example, the volume of the two cones shown is different, even though they have the same area once projected to perspective space.

X. Conclusions

An important line of investigation reported on here is the comparison of the results on synthetic data with those on images. We found that the performance of the algorithms which take advantage of the details of the statistics of the world was worse then expected, based on our experiments with synthetic data. With synthetic data, Color by Correlation and the neural net performed very well. Unfortunately, we were not able to realize this promised performance with our image data set. Instead, exploiting pixel intensity proved to be a larger advantage than exploiting detailed image chromaticity statistics. Thus it is unclear in general when exploiting image chromaticity statistics will work well enough to offset the performance drop due to ignoring pixel intensity. In our synthetic test domain, this was the case; with our image data, it was not.

Exploiting detailed image statistics requires a good match between the statistics used for calibration and the statistics the vision system encounters in the world. Ensuring a good match requires better characterization of image statistics than currently available. Even if the statistics are known, they may not necessarily be exploitable for significant gain, as was the case with our image data set. Thus the potential for improvement in a given application domain remains very much an open question.

A second important finding is that 3-D gamut mapping yields better chromaticity estimates than 2-D gamut mapping. Since 3-D methods have more information (e.g., pixel intensity), they should perform at least as well, but it was unclear whether a worthwhile performance gain should be expected. We found that in every experiment, including all experiments with synthesized data, that the gain was significant.

A third finding is that using the detailed image statistics does help when they are known. This was the case in the experiments with synthesized data as well as in the second post-hoc experiment (Table IV). Thus our work leads directly to the suggestion of combining the strengths of 3-D methods with those which utilize detailed image statistics. We see this, together with improved understanding of image statistics, as the most promising direction for improving computational color constancy performance.

References

[1] Data for computer vision and computational color vision. [Online]. Available: http://www.cs.sfu.ca/~color/data.
[2] Implementations for selected color constancy algorithms. [Online]. Available: http://www.cs.sfu.ca/~color/code.
[3] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms—Part I: Methodology and experiments with synthesized data," *IEEE Trans. Image Processing*, vol. 11, pp. 972–984, Sept. 2002.
[4] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.
[5] B. V. Funt, M. S. Drew, and J. Ho, "Color constancy from mutual reflection," *Int. J. Comput. Vis.*, vol. 6, pp. 5–24, 1991.
[6] B. V. Funt and M. S. Drew, "Color space analysis of mutual illumination," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1319–1326, Dec. 1993.
[7] B. K. P. Horn, "Determining lightness from an image," *Comput. Vis., Graph., Image Process.*, vol. 3, pp. 277–299, 1974.
[8] A. Blake, "Boundary conditions for lightness computation in Mondrian world," *Comput. Vis., Graph., Image Process.*, vol. 32, pp. 314–327, 1985.
[9] B. V. Funt, M. S. Drew, and M. Brockington, "Recovering shading from color images," in *Proc. 2nd Eur. Conf. Computer Vision*, 1992, pp. 124–132.

[10] K. Barnard, G. Finlayson, and B. Funt, "Color constancy for scenes with varying illumination," *Comput. Vis. Image Understand.*, vol. 65, pp. 311–321, 1997.
[11] K. Barnard and B. Funt, "Camera characterization for color research," *Color Res. Applicat.*, vol. 27, no. 3, pp. 152–163, 2002.
[12] B. Funt and V. C. Cardei, "Bootstrapping color constancy," in *Proc. Human Vision and Electronic Imaging IV*, San Jose, CA, 1999, pp. 421–428.
[13] P. L. Vora, J. E. Farrell, J. D. Tietz, and D. H. Brainard, "Image capture: Simulation of sensor responses from hyperspectral images," *IEEE Trans. Image Processing*, vol. 10, pp. 307–316, Feb. 2001.
[14] ——, "Digital color cameras. 1. Response models," Hewlett-Packard Lab., Tech. Rep. HPL-97-53, [Online]. Available: http://www.hpl.hp.com/techreports/97/HPL-97-53.html., 1997.
[15] J. A. Worthey, "Limitations of color constancy," *J. Opt. Soc. Amer. [Suppl.]*, vol. 2, pp. 1014–1026, 1985.
[16] 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.
[17] 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.
[18] G. D. Finlayson, "Coefficient color constancy," Ph.D. dissertation, Sch. Comput., Simon Fraser Univ., Burnaby, BC, Canada, 1995.
[19] 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.
[20] K. Barnard, F. Ciurea, and B. Funt, "Sensor sharpening for computational color constancy," *J. Opt. Soc. Amer. A*, vol. 18, pp. 2728–2743, 2001.
[21] J. J. McCann, S. P. McKee, and T. H. Taylor, "Quantitative studies in Retinex theory," *Vis. Res.*, vol. 16, pp. 445–458, 1976.
[22] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, pp. 108–129, 1977.
[23] 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.
[24] D. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, pp. 5–36, 1990.
[25] K. Barnard, "Computational color constancy: Taking theory into practice," M.Sc. thesis, Sch. Comput. Sci., Simon Fraser Univ., Burnaby, BC, Canada, 1995, [Online]. Available: ftp://fas.sfu.ca/pub/cs/theses/1995/KobusBarnardMSc.ps.gz.
[26] G. D. Finlayson, "Color in perspective," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 1034–1038, 1996.
[27] G. Finlayson and S. Hordley, "A theory of selection for gamut mapping color constancy," *Image Vis. Comput.*, vol. 17, pp. 545–588, 1999.
[28] K. Barnard, "Practical color constancy," Ph.D. dissertation, Sch. Comput. Sci., Simon Fraser Univ., Burnaby, BC, Canada, 1999, [Online]. Available: ftp://fas.sfu.ca/pub/cs/theses/1999/KobusBarnardPhD.ps.gz.
[29] 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.
[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, 2001.
[31] P. M. Hubel and G. Finlayson, "White point estimation using correlation matrix memory," U.S. Patent 6 038 339, 2000.
[32] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *J. Opt. Soc. Amer. A*, vol. 14, pp. 1393–1411, 1997.
[33] 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.
[34] B. V. Funt, V. C. Cardei, and K. Barnard, "Method of estimating chromaticity of illumination using neural networks," U.S. Patent 5 907 629, 1999.
[35] V. Cardei. (2000) A neural network approach to color constancy. Simon Fraser Univ., Burnaby, BC, Canada. [Online]. Available: http://www.cs.sfu.ca/~color/publications/VCardeiPhD/VCardei_PhD_Thesis.pdf
[36] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Res. Applicat.*, no. 3, pp. 147–151, 2002.
[37] S. Tominaga, A. Ishida, and B. A. Wandell, "Further research on the sensor correlation method for scene illuminant classification," in *Proc. 8th Color Imaging Conf.*, Scottsdale, AZ, 2000, pp. 189–194.
[38] S. Tominaga, S. Ebisui, and B. A. Wandell, "Color temperature estimation of scene illumination," in *Proc. IS&T/SID 7th Color Imaging Conf.: Color Science, Systems, and Applications*, Scottsdale, AZ, 1999, pp. 42–47.
[39] 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.

Kobus Barnard received the Ph.D. degree in computer science from Simon Fraser University, Burnaby, BC, Canada, where he specialized in computational color constancy.

He is a Postdoctoral Fellow in computer vision at the University of California at Berkeley. His current research projects include using large annotated image datasets for minimally supervised learning of image semantics, and the application of computer vision to the organization and effective use of large image collections. He also continues to work on understanding scene illumination, especially in the case of spatially varying illumination and shadow identification.

Lindsay Martin received the B.Sc. degree in computing science from Simon Fraser University, Burnaby, BC, Canada, in 1996. He worked as a Research Assistant under the supervision of Dr. B. Funt.

He is currently employed as a Software Developer with InTime Solutions, Inc.

Adam Coath is pursuing the M.S. degree in computing science at Simon Fraser University, Burnaby, BC, Canada.

His current research focuses on information dissemination in communication networks.

Brian Funt received the Ph.D. degree in computer science in 1976 from the University of British Columbia, Vancouver, BC, Canada.

He is Professor of computing science at Simon Fraser University, Burnaby, BC, where he has worked since 1980. Since the early 1980s, his research has focused on computational models of color with a special emphasis on color constancy.