Bayesian Color Constancy for Outdoor Object Recognition

Yanghai Tsin[†] Robert T. Collins[†] Visvanathan Ramesh[‡] Takeo Kanade[†]

[†] The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{ytsin,rcollins,tk}@cs.cmu.edu

[‡]Imaging & Visualization Department
Siemens Corporate Research
Princeton, NJ 08540
rameshv@scr.siemens.com

Abstract

Outdoor scene classification is challenging due to irregular geometry, uncontrolled illumination, and noisy reflectance distributions. This paper discusses a Bayesian approach to classifying a color image of an outdoor scene. A likelihood model factors in the physics of the image formation process, the sensor noise distribution, and prior distributions over geometry, material types, and illuminant spectrum parameters. These prior distributions are learned through a training process that uses color observations of planar scene patches over time. An iterative linear algorithm estimates the maximum likelihood reflectance, spectrum, geometry, and object class labels for a new image. Experiments on images taken by outdoor surveillance cameras classify known material types and shadow regions correctly, and flag as outliers material types that were not seen previously.

1. Introduction

Color is an important feature for many vision tasks such as segmentation[9], object recognition[2] and image retrieval[20]. However, the apparent color of a surface varies with illumination, and it is necessary to account for this apparent color change to use color robustly. Color constancy [6, 1, 5] is one way to deal with this problem. Color constancy algorithms attempt to estimate the illuminant spectrum and compensate for its contribution to image appearance. Color constancy would seem to be an appealing preprocessing step for color-based vision tasks. Unfortunately, previous work shows that it is not sufficient to merely concatenate a color constancy algorithm with an object recognition algorithm [8, 20].

Many interesting vision applications, such as surveillance and robot navigation, involve irregular geometry, uncontrolled lighting and random reflectance distributions. These real world applications seriously challenge existing color constancy algorithms, which have been tested only in synthetic and laboratory settings. There exist very few color constancy algorithms that work on real images[6, 2].

We have observed several factors that contribute to the gap between color constancy theory and applications. First, color constancy algorithms deal with very general cases. Many color constancy algorithms assume the existence of statistical invariants governing the natural world. For example, the gray-world method assumes known mean reflectance of any natural scene, and there is work attempting to estimate such statistics from large sets of images [17]. We seek to avoid such global assumptions by learning reflectance distributions only for classes of objects observed in a set of training images, leading to a specific and well-defined estimation problem.

Second, existing color imaging models have unpredictable accuracy. For example, the diagonal transformation model [6, 8] gives good approximations only when the camera has narrow and non-overlapping spectral sensitivity functions. The generalized diagonal transform [5] performs better than the diagonal transformation model, but assumes very low dimensionality of the reflectance and spectrum. Finite dimensional linear models [12, 1, 22] require explicit use of illumination basis functions and camera sensitivity functions, which are not always available and accurate. Our observation is that only the coefficients used to combine basis functions are important for color constancy. As a result, we propose a color imaging model that has the simple bilinear form of the diagonal model, yet without the bias introduced by inaccurate sensitivity and basis functions.

Finally, the computational burden of solving a color constancy problem is non-trivial when a high degree non-linear optimization problem is imposed on each pixel[1]. We introduce an iterative linear update method that reduces the computational cost dramatically, thus making it possible to work on real images.

By appropriately dealing with the above issues, we have developed a color-based object recognition algorithm that can be applied directly to real world environments. In this work we take advantage of the low dimensionality of outdoor light spectra, although indoor light spectra can be treated in a similar manner.

We denote scalars as normal font characters, such as g. Vectors are denoted as bold lowercase characters (e.g. $\mathbf{v}$), matrices as bold uppercase (e.g. $\mathbf{M}$). Random variables and estimates are denoted with a hat (e.g. $\hat{a}$).

2. Color Image Formation Model

We study Lambertian surfaces in this paper. For a Lambertian surface, the measured intensity ρ_c of channel c, $c = 1, 2, \ldots, n_c$, is

$$\rho_c = g \int_\lambda f_c(\lambda) s(\lambda) l(\lambda) d\lambda$$

Here g is an "effective light intensity" determined by the scene geometry and the absolute light intensity (see Section 3); $f_c(\lambda)$ is the sensitivity function of channel c; $s(\lambda)$ denotes the reflectance; $l(\lambda)$ represents the *normalized* light spectrum (chromaticity of a light source); and all the above variables are functions of the wavelength λ. Since RGB color cameras are the most commonly used sensors in vision research, without loss of generality we assume $n_c = 3$ hereafter.

By discretizing the reflectance s, light spectrum l and sensitivity function f_c into N samples, and denoting each discretized function as a column vector, we obtain an equivalent vector representation

$$\rho_c = g \boldsymbol{l}^T \boldsymbol{D}(\boldsymbol{f_c}) \boldsymbol{s} \tag{1}$$

where $\boldsymbol{D}(\boldsymbol{f_c})$ is the $N \times N$ diagonal matrix with $\boldsymbol{f_c}$ as diagonal elements.

We adopt finite dimensional linear models for both reflectance [13, 16, 3] and illuminant spectrum[11, 18]. Assume the reflectance and spectrum are spanned by the column spaces of the matrices $\boldsymbol{B_s}$ and $\boldsymbol{B_l}$ respectively. The reflectance and spectrum can be rewritten as

$$\begin{cases} \boldsymbol{s} = \boldsymbol{B_s} \boldsymbol{\alpha} \\ \boldsymbol{l} = \boldsymbol{B_l} \boldsymbol{\beta} \end{cases} \tag{2}$$

Here $\boldsymbol{\alpha} \in \mathcal{R}^{n_\alpha}$ and $\boldsymbol{\beta} \in \mathcal{R}^{n_l}$ are coefficient vectors with much lower dimensionality than N. Previous research shows that both the natural light spectrum [11, 18] and reflectance [13, 16, 3] can be approximated accurately with such low dimensional linear systems.

Substituting (2) into (1) we get

$$\rho_c = g \boldsymbol{\beta}^T \boldsymbol{B_l}^T \boldsymbol{D}(\boldsymbol{f_c}) \boldsymbol{B_s} \boldsymbol{\alpha}$$

Denoting the n_l-element vector $\boldsymbol{B_l}^T \boldsymbol{D}(\boldsymbol{f_c}) \boldsymbol{B_s} \boldsymbol{\alpha}$ as $\boldsymbol{\sigma}_c$, and assuming that g is known, we have a simple bilinear model for color image formation

$$\rho_c = g \boldsymbol{\beta}^T \boldsymbol{\sigma}_c. \tag{3}$$

Here, $\boldsymbol{\sigma}_c$ can be considered as a filtered version of the original reflectance, where the filter is determined by the basis functions and the sensitivity functions.

Finally, for brevity we denote $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3)^T \in \mathcal{R}^3$ as the color vector and $\boldsymbol{S} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3) \in \mathcal{R}^{n_l \times 3}$ as the reflectance matrix. We then have one big matrix equation

$$\boldsymbol{\rho} = g \boldsymbol{S}^T \boldsymbol{\beta} \quad . \tag{4}$$

In Section 5 and Appendix A we will encounter the problem of estimating reflectance $\boldsymbol{\sigma}_c$, for $c = 1, 2, 3$. It is more convenient to represent reflectance as a vector $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1^T, \boldsymbol{\sigma}_2^T, \boldsymbol{\sigma}_3^T)^T \in \mathcal{R}^{3n_l}$ than as the matrix $\boldsymbol{S}$. By writing a lighting matrix $\begin{bmatrix} \boldsymbol{\beta} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\beta} \end{bmatrix} \in \mathcal{R}^{3n_l \times 3}$ we have the dual form of (4)

$$\boldsymbol{\rho} = g \boldsymbol{B}^T \boldsymbol{\sigma} \quad . \tag{5}$$

The bilinear relationship (3) has long been observed in the computer vision literature [1, 23, 22]. The important difference here is that our representation is independent of the basis function selection. In [1, 22] light spectrum basis functions are obtained by extracting principle components from a large set of light samples measured by spectroradiometers, and camera sensitivity functions are selected heuristically, for example to approximate the human eye response [19]. Such choices work fine for simulations. However, real applications of this approach require accurate specification of basis and sensitivity functions. Our model (3), on the other hand, discards such error-prone procedures, and therefore can be used with any camera. The concept of a lighting matrix in (4) and reflectance matrix in (5) also has been previously used in computer vision [5]. Our contribution is that we parameterize them using $4n_l$ variables in a manner that is independent of basis functions and sensitivity functions.

3. Statistical Models

To model the variations in the reflectance $\boldsymbol{\sigma}$, illuminant spectrum $\boldsymbol{\beta}$, and effective light intensity g (due to geometric attenuation), we treat them as random variables.

First of all, natural objects exhibit randomness in their apparent color. We model this randomness of color pigment distribution using the conditional probability distribution $p(\boldsymbol{\sigma}|o)$ where o stands for the object class. For simplicity we consider single colored objects only, and thus use a unimodal distribution for $p(\boldsymbol{\sigma}|o)$. Multi-colored objects could be modeled by extension as multi-modal mixtures of single-mode distributions.

Secondly, when the light sources (or weather conditions, represented by the symbol w) are known, the illuminant spectrum is predictable. This distribution of the light spectrum given the light source is modelled as $p(\boldsymbol{\beta}|w)$.

And finally, the irregularity of the scene geometry contributes to the randomness of the effective light strength g. The effective light strength at a scene point is a function of the incident light strength I and the lighting geometry at the point. Specifically, we model the sunlight (with incident strength I_{sun}) as a point light source and the skylight (with incident strength I_{sky}) as an illumination dome. The contribution of the geometry to the point light source is an attenuation factor $a = cos\theta$, with θ representing the angle subtended by the irradiance direction and the surface normal. In an open space the skylight is attenuated by a factor of $b = \frac{1+\cos\phi}{2}$, with ϕ being the angle subtended by the surface normal and the vertical direction[10]. Thus $g = aI_{sun}$ and $g = bI_{sky}$ for the two cases respectively. When occlusion occurs in the scene, g will be further attenuated. For simplicity we assume $a \cdot b = 0$, which means that only one light source will be dominant at a surface patch. A more elaborate model can be adopted for additional accuracy, such as the modelling of g as a linear combination of the two light sources at each scene point.

When a site model and ephemeris data (latitude, longitude, date and time) is available, g can be computed for each scene patch up to a constant scale, which is due to the unknown absolute light strength. For many cases, however, a 3D scene model will not be available. Instead, we assume a piece-wise near-planar scene and model $g(x)$ as a single mode random variable for each local patch, with the randomness introduced by the small fluctuation of scene geometry within the patch. The distribution of g is conditioned on the light source because g is a scaled version of I_{sun} or I_{sky}. We write this distribution as $p(g|w)$.

It can be noted that the restrictions imposed by the probabilistic models for g essentially translate to constraints on the effective light intensity change. It enables us to tell surfaces with the same chromaticity but different brightness from each other. To illustrate the point, consider an image of a white road mark (with color vector $k\boldsymbol{\rho}$) on a concrete pavement (with color $\boldsymbol{\rho}$). The part due to road mark can be explained as $(\boldsymbol{\sigma}) \times (kg\boldsymbol{\beta})$ (the same reflectance, different lighting), or $(k\boldsymbol{\sigma}) \times (g\boldsymbol{\beta})$ (the same lighting, different reflectance). The knowledge of g distribution tells us that the latter is more likely, resulting in a higher aposteriori probability that the given data may correspond to a road mark.

4. Learning Statistical Distributions from Multiple Observations

Due to Maloney and Wandell[23, 12], it is now well known that only a $n_c - 1$ dimensional reflectance descriptor can be recovered uniquely from a single n_c channel multi-spectral image. Unfortunately, conventional color CCD cameras have only three channels, and two dimensional descriptors have been shown to be insufficient for color constancy tasks[8]. Following [22, 4] we estimate the reflectance and illuminant spectrum from multiple registered images. Our bilinear color imaging model (3) makes the recovery process simple.

A small planar scene patch containing material types of interest is selected for study. Over time it is uniformly illuminated by lights with different spectrum $\boldsymbol{\beta}$. We denote the effective light strength at time t as $g(t)$. The color vector observed at time t and at pixel x is $\boldsymbol{\rho}(x,t)^T = g(t)\boldsymbol{\beta}(t)^T\boldsymbol{S}(x)$. Assume we observed F frames of P pixels, we can write the color measurement matrix as,

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{\rho}(x_1,t_1)^T & \dots & \boldsymbol{\rho}(x_P,t_1)^T \\ \vdots & \vdots & \vdots \\ \boldsymbol{\rho}(x_1,t_F)^T & \dots & \boldsymbol{\rho}(x_P,t_F)^T \end{bmatrix}$$

By writing the light spectrum over time as a matrix $\boldsymbol{L}$, and reflectance across space as $\boldsymbol{R}$, we get the following:

$$\boldsymbol{L} = \begin{bmatrix} g(t_1)\boldsymbol{\beta}(t_1)^T \\ \vdots \\ g(t_F)\boldsymbol{\beta}(t_F)^T \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} \boldsymbol{\sigma}(x_1) & \dots & \boldsymbol{\sigma}(x_P) \end{bmatrix}$$

and therefore can derive the simple relationship

$$\boldsymbol{M} = \boldsymbol{L}\boldsymbol{R} \tag{6}$$

When the number of observations F is greater than the model dimension n_l, we can recover the reflectance and light spectrum up to a non-singular $n_l \times n_l$ matrix. This recovery can be easily achieved by applying singular value decomposition(SVD) on the color measurement matrix $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}$. The estimate for the reflectance matrix $\boldsymbol{R}$ is given by the first n_l rows of $\boldsymbol{V}$, and the estimate for the light spectrum matrix $\boldsymbol{L}$ is given by the first n_l column of $\boldsymbol{U}\boldsymbol{W}$. Each row of the estimated matrix $\hat{\boldsymbol{L}}$ is normalized to give a estimate of $\boldsymbol{\beta}(t)$.

Each estimated $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\beta}}$ is considered to be a sample from the reflectance and illuminant spectrum, accordingly. If the pixels are labeled by object class and each frame is labeled by light source class, both the reflectance distribution given object class $p(\boldsymbol{\sigma}|o)$ and the spectrum distribution given light source $p(\boldsymbol{\beta}|w)$ can be estimated by sample statistics. If such labels are not available, the samples can be segmented into distribution modes using unsupervised clustering methods such as EM[15], and after clustering each mode can be assigned a meaning.

5. Inference of Scene Contents

Given the estimated distributions for reflectance and lighting spectra, our goal is now to infer scene contents when an image of a novel scene is presented. That is, we want to determine the object class (material type) and light sources

from pixels in the image. Following [1, 7, 17], we formulate the problem within a Bayesian framework.

When there is no noise, (4) gives an exact color prediction. However, our measurement $\hat{\boldsymbol{\rho}}$ is corrupted by Gaussian noise with covariance matrix $\boldsymbol{\Sigma_\rho}$, so the probability of observing an actual color is given by

$$p(\hat{\boldsymbol{\rho}}|\boldsymbol{S},\boldsymbol{\beta},g) = (2\pi|\boldsymbol{\Sigma_\rho}|)^{-3/2}\, exp\{-\|\hat{\boldsymbol{\rho}} - g\boldsymbol{S}^T\boldsymbol{\beta}\|_{\boldsymbol{\Sigma_\rho}}\}$$

Here $|\boldsymbol{A}|$ is the determinant of $\boldsymbol{A}$ and $\|\boldsymbol{v}\|_{\boldsymbol{\Sigma}}$ is the Mahalanobis distance $\boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v}$. This is the generative model of observing a color vector.

Inference of scene contents proceeds in the other direction by determination of the *maximum a posteriori* (MAP) estimate of the scene contents given the observed color vector $\hat{\boldsymbol{\rho}}$,

$$[\hat{o},\hat{w},\hat{\boldsymbol{S}},\hat{\boldsymbol{\beta}},\hat{g}] = \underset{[o,w,\boldsymbol{S},\boldsymbol{\beta},g]}{argmax}\ p(o,w,\boldsymbol{S},\boldsymbol{\beta},g|\hat{\boldsymbol{\rho}}) \qquad (7)$$

Applying Bayes rule, it is easy to show that

$$p(o,w,\boldsymbol{S},\boldsymbol{\beta},g|\hat{\boldsymbol{\rho}}) \propto$$
$$p(\hat{\boldsymbol{\rho}}|\boldsymbol{S},\boldsymbol{\beta},g)p(\boldsymbol{\beta}|w)p(\boldsymbol{S}|o)p(g|w)p(w)p(o) \qquad (8)$$

Here we have assumed conditional independence of $\hat{\boldsymbol{\rho}}$ with w and o given $\boldsymbol{S}$, $\boldsymbol{\beta}$ and g, and we also assume independence of the reflectance and light spectrum.

Conceptually, we have a hierarchical prior model and a sensor likelihood model in (8). At the highest level, $p(w)$ and $p(o)$ define the prior probability of observing light source w and object o in a given scene. If $p(w)$ is available (by knowing the time of day, for example), and if $p(o)$ is available (when working in a familiar scene), the prediction of scene contents can be greatly improved. The prior densities $p(\boldsymbol{\beta}|w)$, $p(\boldsymbol{S}|o)$, and $p(g|w)$ represent prior knowledge of the light spectrum, reflectance and geometry, respectively. Without this knowledge it is not possible to recover a $3 \times n_l$ dimensional reflectance, an n_l dimensional spectrum and the three scalars g,o and w from only a 3 dimensional observation $\hat{\boldsymbol{\rho}}$. Finally, the likelihood model for the measurements is derived from the physics of the image formation process and the sensor error model, where the color formation process $p(\hat{\boldsymbol{\rho}}|\boldsymbol{S},\boldsymbol{\beta},g)$ is described. Bayes rule (8) provides us with a scientific way for integrating information among these different levels of knowledge.

Solving the MAP problem (7) is not trivial. For each pixel it involves a nonlinear optimization problem with $3 \times n_l + 1$ continuous variables ($\boldsymbol{S}$ and g) and two discrete variables (w and o). Furthermore, globally, there is also one n_l dimensional variable ($\boldsymbol{\beta}$) to estimate for each light source. In response to this problem, we have developed an iterative solution method based on linear updating. Starting from some initialization, the algorithm iteratively updates reflectance, geometry, light spectrum and classifies each pixel. A detailed mathematical derivation for Gaussian statistical models is presented in Appendix A, and we present our algorithm in Section 6. At each iteration the *a posteriori* probability increases, and the method is guaranteed to converge to a local maximum. One important feature of our iterative linear method is that in each step of updating the reflectance, we transform inversion of $3n_l \times 3n_l$ matrices into inversion of 3×3 matrices. For our experiments we have chosen $n_l = 10$. Considering that the matrix inversion is performed for each pixel, and that it is an $o(n^3)$ operation, use of our linear method greatly reduces the computational cost.

6. Algorithm for Solving the MAP

After the learning process we have statistical knowledge of both the reflectance and the lighting. We utilize this knowledge to compute the *mean color chart*, the table of typical colors of different material types under different lighting conditions. The "typical" color of a surface o under light source w is given by $\bar{\boldsymbol{\rho}}(o,w) = \boldsymbol{\mu_{\sigma|o}}\boldsymbol{\mu_{\beta|w}}$, with $\boldsymbol{\mu_{\sigma|o}}$ and $\boldsymbol{\mu_{\beta|w}}$ representing the mean reflectance of o and mean spectrum of w respectively.

We initialize our algorithm by comparing the observed color vector with each of the mean colors $\bar{\boldsymbol{\rho}}(o,w)$. A pixel belongs to object class o and is lit by light source w only if it has similar chromaticity to the mean color $\bar{\boldsymbol{\rho}}(o,w)$. This similarity is measured by the angle subtended by the observed color vector and the mean color $\bar{\boldsymbol{\rho}}(o,w)$. The smaller the angle, the more similar the chromaticity.

We define a local hypothesis at a pixel as the set of estimates $H = \{\hat{\boldsymbol{\sigma}},\hat{g},\hat{o},\hat{w}\}$. Our goal is to find the best hypothesis that maximizes the MAP (8). To avoid mistakes introduced by the initialization step we keep n_h hypotheses $H_1,\ldots,H_{n_h}$ for each pixel, sorted by descending likelihoods of the hypotheses. Even when the best initial hypothesis H_1 is wrong, the correct hypothesis can still be included in the population and can emerge as the best hypothesis during the estimation iterations.

The initialization algorithm is summarized as following,

- *For each pixel, whose observed color is* $\hat{\boldsymbol{\rho}}$
  - *Compute the angles* $\theta(o,w)$ *subtended by* $\hat{\boldsymbol{\rho}}$ *and each of the mean colors* $\bar{\boldsymbol{\rho}}(o,w)$.
  - *Sort* $\theta(o,w)$ *in ascending order.*
  - *For the* i*th smallest angle* $\theta(o,w)$*, initialize* H_i *as follow*
    - *Set* $\hat{o}$ *and* $\hat{w}$ *to the corresponding* o *and* w
    - *Set* $\hat{\boldsymbol{\sigma}}$ *to be* $\boldsymbol{\mu_{\sigma|o}}$.
    - *Set* $\hat{g}$ *to be* $\frac{\hat{\boldsymbol{\rho}}^T\bar{\boldsymbol{\rho}}(o,w)}{\|\bar{\boldsymbol{\rho}}(o,w)\|}$.
- *For each* w*, form* $\mathcal{N}_w$*, the set of pixels whose best hypothesis predicts* w *as the light source.*

- *Estimate parameters of the distribution $p(g|w)$ from the set of samples $\mathcal{G}_w = \{g(x)|x \in \mathcal{N}_w\}$.*
- *Set the initial spectrum of class w as $\boldsymbol{\mu}_{\beta|w}$.*

We are exploring the joint space of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, g, o and w for the best hypothesis. Unlike the continuous variables the discrete variables o and w cannot be updated analytically. The complete brute-force method for solving the MAP problem retains hypotheses corresponding to each combination of the discrete variables $\hat{o}$ and $\hat{w}$. The probability for each hypothesis is computed and every hypothesis is updated until convergence, when the hypothesis with the best probability is chosen as the result. This procedure becomes increasingly costly as the cardinality of $\hat{o}$ and $\hat{w}$ increase. By utilizing the mean color chart comparison criterion we discard many quite unlikely hypotheses from the start, thus increasing the algorithm's efficiency.

After initialization, we can refine the MAP estimation iteratively.

- *Divide the image into overlapping windows (assume each window corresponds to a near-planar surface of the scene).*
- *For each window*
  - *Form $\mathcal{N}_w$, the set of pixels whose best hypothesis predicts w as the light source.*
  - *Estimate parameters of the distribution $p(g|w)$ from the set of samples $\mathcal{G}_w = \{g(x)|x \in \mathcal{N}_w\}$.*
  - *For each pixel,*
    - *For each hypothesis H_i, $i = 1, \ldots, h_h$,*
      - *Update $\hat{\boldsymbol{\sigma}}$ according to (11)-(14).*
      - *Update $\hat{g}$ according to (16).*
- *For each pixel*
  - *Compute likelihood of the hypotheses according to (8).*
- *Update light spectrum of each class w according to (18)-(21).*
- *If $n_h > 1$, $n_h - 1 \rightarrow n_h$, delete the least likely hypothesis of each pixel.*
- *Iterate until convergence.*

7. Experiments

7.1. Data Collection

We have collected experimental data from a static surveillance camera mounted on a building roof. The acquisition hardware is a Sony EVI-330 color video camera and a Matrox Meteor II. Fourteen images were collected every five minutes over a period of two days. Each of the fourteen images were taken in quick succession, at different exposures. From this data the camera response function was calibrated and high dynamic range (HDR) images together with estimates of the variance at each pixel were computed using our calibration algorithm [21]. The result is essentially a measurement of scene radiance $\hat{\boldsymbol{\rho}}$ and its uncertainty $\boldsymbol{\Sigma}_{\boldsymbol{\rho}}$, for each pixel.

7.2. Learning

An approximately planar training image is shown in Figure 1(a). Twenty-one such registered and uniformly illuminated images were used to train the algorithm. Each patch contains two material types of interest: vegetation and road pavement, and the pixels corresponding to these two material types were manually labeled. Reflectance at each of the selected pixels and the illuminant spectrum for each frame were then estimated using the method introduced in Section 4. To evaluate how many basis functions are needed, we evaluated the model fitting error $\boldsymbol{M - UWV}$ while varying the number of basis functions. Figure 2(a) shows the median of the relative error plotted against the number of basis functions. Note that the numbers are an estimate of the standard deviation of the possible relative errors[14]. If the errors obey a Gaussian distribution, it is not uncommon to have errors as much as four times the plotted number. For our experiments we chose 10 basis functions. A scatter plot of the first three principal components of the estimated reflectance is shown in Figure 2(b). We see that the reflectance corresponding to the two classes are nicely separated, and that it is reasonable to fit a Gaussian model to each of them.

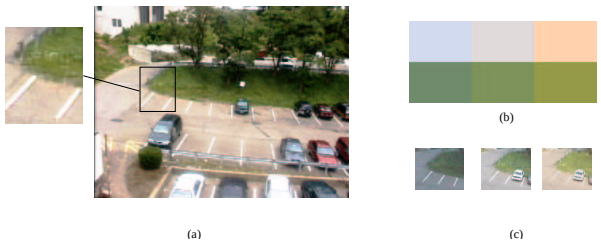


Figure 1: *(a) A sample input image and the small scene patch selected for training the algorithm. (b) The mean color chart. The first row represents colors of the road pavement under different light source, and the second row represents mean colors of the vegetation. The three light sources (from left to right) are: early morning, shadow, and sunlight. (c) Original image patches under the three light sources. Notice the similarity between the image patch and the corresponding mean color, and the obvious color difference under different light sources.*

After estimating the reflectance distributions, we added 61 other patches (from the same location) into consideration. In these images non-uniform illumination was allowed. Given the known reflectance, we estimated one light spectrum sample from each 5×5 window in the image via least squares. These light samples lie on a small area of a 10 dimensional hypersphere (because we specified to use 10 basis functions). We applied the EM clustering algorithm to cluster these samples into three classes, and fit a Gaussian model to each mode. The first

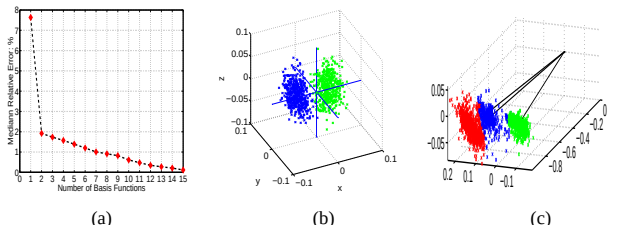


Figure 2: *(a) The relative median error as a function of the number of basis functions. (b) Scatter plot for the two surface types of interest. (c) Illuminant spectrum clusters*

three principal components of the light spectrum are shown in Figure 2(c). We are not claiming that there are only three modes describing all possible daylight spectra. However, these three explain our data very well. The three modes correspond to the light spectra of early morning, shadow and sunlight. A computed mean color chart is shown in Figure 1(b). In Figure 1(c) we show sample images taken from the original data. Image brightness has been adjusted so that only the chromaticity matters. Notice the similarity between the mean color chart and the real images, and the distinct color changes under different light sources. (Please visit http://www.cs.cmu.edu/∼ytsin/research/bayesiancc/ for color images).

7.3. Inference

After training, the estimated statistical models were used for classification. Only HDR images not included in the training images were used for testing.

We applied the initialization algorithm in Section 6 to the data. The initial classification found by the initialization procedures is shown in Figure 3. It is interesting to notice that this simple and low-cost initialization method generates a fairly good segmentation according to material types and light sources. The results suggest that the cached mean color information for a familiar scene can provide important information for image understanding. But at this early stage we can not reliably detect outliers because of initialization errors.

We further applied the iterative updating algorithm in Section 6. The results after convergence are shown in Figure 4. The algorithm has correctly identified many regions corresponding to vegetation and road pavement. By manually setting a threshold on the *a posteriori* probability, we see that it is possible to successfully detect outliers as well, such as parked vehicles, painted road marks, tree trunks and buildings.

We compared the computational cost of our iterative linear method with that of the Levenberg-Marquardt (LM) algorithm. For each combination of o and w it takes about 1 second for the LM algorithm to update one pixel (31 dimensional continuous variables). For a 320x240 image it would have taken days for it to converge. Our linear iterative algorithm takes several minutes to converge on the whole image using the same computer.

A robustness study is a topic we are currently pursuing. Initial tests on other images show repeatable classification results.

8. Conclusion

To the best of our knowledge, there has not been a previous color constancy algorithm that is applicable in an outdoor, uncontrolled environment. By learning customized surface reflectance and lighting distributions, we have successfully combined a color constancy algorithm with an object recognition algorithm and have applied them in outdoor scenes. The approach is based on statistical learning and inference. A Bayesian estimation scheme is presented wherein the prior scene knowledge, i.e. lighting, object/material classes, and geometry, is integrated with a likelihood model motivated from the physics of image formation and a sensor error model. The experimental results confirm the validity of our model assumptions in the outdoor scenario tested.

We have adopted the Gaussian noise model in our experiments due to its computational simplicity. However, our algorithm is not limited to Gaussian models, or even to single-mode distributions. When the Gaussian assumption is no longer valid, the learning and inference methods in Section 4 and 5 still hold. The solution of the MAP problem may become much different, however. In the most computationally challenging cases, general sampling and resampling techniques [7] can still be applied to achieve a solution.

A. Solving the MAP Problem

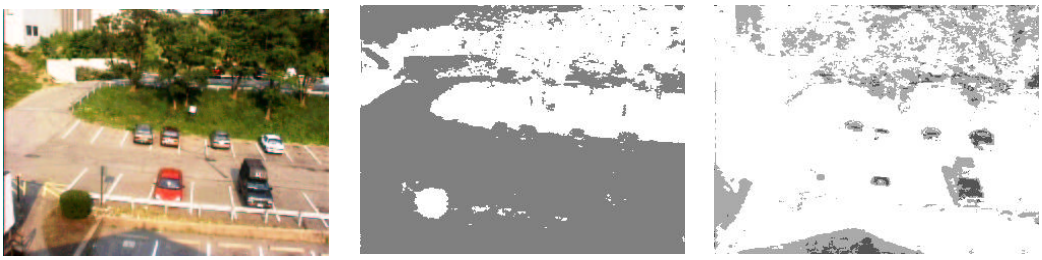
We discuss how to update a hypothesis based on estimates available at step n, given Gaussian statistical models for reflectance, spectrum and geometry. First, we discuss how to update reflectance. We assume all other variables are known and are equal to $\hat{o}^{(n)}$, $\hat{w}^{(n)}$, $\hat{g}^{(n)}$ and $\hat{\boldsymbol{\beta}}^{(n)}$. The cost function to be minimized is

$$\mathcal{COST}_{\boldsymbol{\sigma}} = \|\hat{\boldsymbol{\rho}} - \hat{g}^{(n)}\hat{\boldsymbol{B}}^{(n)T}\boldsymbol{\sigma}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\rho}}} + \|\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}|o}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}} \quad (9)$$

Here $\hat{\boldsymbol{B}}^{(n)}$ is the estimate of the lighting matrix $\boldsymbol{B}$ at step n. $\boldsymbol{\mu}_{\boldsymbol{\sigma}|o}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}$ are the mean and covariance of the reflectance of object class o. The solution to (9) is,

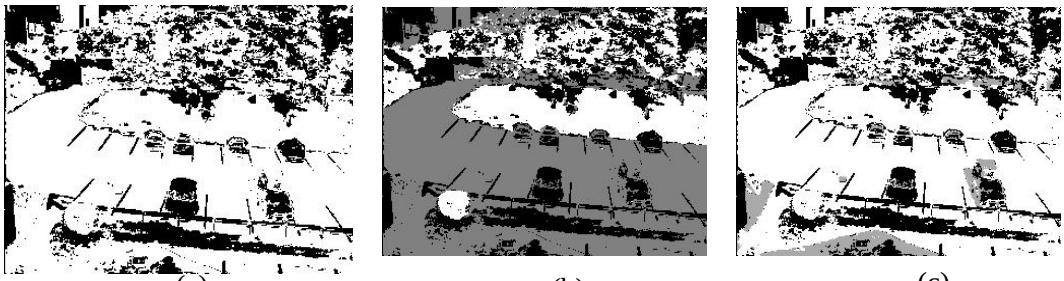
$$\begin{aligned}\hat{\boldsymbol{\sigma}}^{(n+1)} &= \left(\hat{g}^{(n)2}\hat{\boldsymbol{B}}^{(n)}\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}\hat{\boldsymbol{B}}^{(n)T} + \boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}^{-1}\right)^{-1} \\ &\quad \left(\hat{g}^{(n)}\hat{\boldsymbol{B}}^{(n)}\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}\hat{\boldsymbol{\rho}} + \boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}^{-1}\boldsymbol{\mu}_{\boldsymbol{\sigma}|o}\right) \quad (10)\end{aligned}$$

The first term on the right hand side of the above equation involves inversion of a $3n_l \times 3n_l$ matrix for each pixel,



(a) (b) (c)

Figure 3: *Initial segmentation based on comparison with the mean color chart (a) The test image. (b) Initial segmentation by material types. White: vegetation. Gray: road pavement. (c) Initial segmentation by light source. White: sunlight. Gray: shadow. Dark gray: early morning*



(a) (b) (c)

Figure 4: *Final results (a) The detected outliers. Outliers are shown as black pixels (b) Final segmentation by material types. White: vegetation. Gray: road pavement. (c) Final segmentation by light source. White: sunlight. Gray: shadow.*

which is very costly. To avoid this high computational burden, we apply the inversion formula

$$(\boldsymbol{A}+\boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{C}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{DA}^{-1}$$

Using this formula, solution (10) is seen to be equivalent to

$$\hat{\boldsymbol{\sigma}}^{(n+1)} = \boldsymbol{K}_1\boldsymbol{\mu}_{\boldsymbol{\sigma}|o} + \boldsymbol{K}_2\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}\hat{\boldsymbol{\rho}} \quad (11)$$

$$\boldsymbol{K}_1 = \boldsymbol{I} - \boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}\hat{\boldsymbol{B}}^{(n)}\boldsymbol{K}_3\hat{\boldsymbol{B}}^{(n)T} \quad (12)$$

$$\boldsymbol{K}_2 = \hat{g}^{(n)}\boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}\hat{\boldsymbol{B}}^{(n)}\left(\boldsymbol{I} - \boldsymbol{K}_3\hat{\boldsymbol{B}}^{(n)T}\boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}\hat{\boldsymbol{B}}^{(n)T}\right) \quad (13)$$

$$\boldsymbol{K}_3 = \left(\frac{\boldsymbol{\Sigma}_{\boldsymbol{\rho}}}{\hat{g}^{(n)2}} + \hat{\boldsymbol{B}}^{(n)T}\boldsymbol{\Sigma}_{\boldsymbol{\sigma}|o}\hat{\boldsymbol{B}}^{(n)}\right)^{-1} \quad (14)$$

which requires inversion of a 3×3 matrix only.

Similarly, the cost function to be minimized for estimating g is

$$\mathcal{COST}_g = \|\hat{\boldsymbol{\rho}} - g\hat{\boldsymbol{S}}^{(n+1)T}\hat{\boldsymbol{\beta}}^{(n)}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\rho}}} + \|g - \hat{\mu}_{g|w}\|_{\hat{\sigma}_{g|w}} \quad (15)$$

Here $\hat{\boldsymbol{S}}^{(n+1)}$ is the estimate of reflectance matrix $\boldsymbol{S}$ at step $n+1$. $\hat{\mu}_g$ and $\hat{\sigma}_g$ is the estimated mean and standard deviation of the effective light intensity for light source w. The solution to (15) is,

$$\hat{g}^{(n+1)} = \frac{\mu_{g|w} + \hat{\sigma}_{g|w}^{(n)2}\hat{\boldsymbol{\rho}}^T\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}\hat{\boldsymbol{S}}^{(n+1)T}\hat{\boldsymbol{\beta}}^{(n)}}{1 + \hat{\sigma}_{g|w}^{(n)2}\hat{\boldsymbol{\beta}}^{(n)T}\hat{\boldsymbol{S}}^{(n+1)}\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}\hat{\boldsymbol{S}}^{(n+1)T}\hat{\boldsymbol{\beta}}^{(n)}} \quad (16)$$

In contrast to the reflectance and effective light strength, the light spectrum is a global variable. The cost function is defined by all the pixels $x \in \mathcal{N}_w$ lit by light source w.

$$\mathcal{COST}_{\boldsymbol{\beta}} = \sum_{x\in\mathcal{N}_w}\Big(\|\hat{\boldsymbol{\rho}}(x) - \hat{g}^{(n+1)}(x)\hat{\boldsymbol{S}}^{(n+1)T}(x)\boldsymbol{\beta}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\rho}}(x)} + \|\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|w}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\beta}|w}}\Big) \quad (17)$$

Here $\hat{\boldsymbol{\rho}}(x)$ is the color vector observed at point x. $\hat{g}^{(n+1)}(x)$ and $\hat{\boldsymbol{S}}^{(n+1)}(x)$ are similarly defined. $\boldsymbol{\mu}_{\boldsymbol{\beta}|w}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}|w}$ are mean and covariance matrix for spectrum $\boldsymbol{\beta}$ of light source class w. For each pixel we have the following system of normal equations

$$\boldsymbol{h}(x)\boldsymbol{\beta} = \boldsymbol{b}(x) \quad (18)$$

$$\boldsymbol{h}(x) = \boldsymbol{\Sigma}_{\boldsymbol{\beta}|w}^{-1} + \hat{g}^{(n+1)2}(x)\hat{\boldsymbol{S}}^{(n+1)}(x)\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}(x)\hat{\boldsymbol{S}}^{(n+1)T}(x) \quad (19)$$

$$\boldsymbol{b}(x) = \boldsymbol{\Sigma}_{\boldsymbol{\beta}|w}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|w} + \hat{g}^{(n+1)}(x)\hat{\boldsymbol{S}}^{(n+1)}(x)\boldsymbol{\Sigma}_{\boldsymbol{\rho}}^{-1}(x)\hat{\boldsymbol{\rho}}(x) \quad (20)$$

The optimal solution is given by

$$\hat{\boldsymbol{\beta}}^{(n+1)} = \left(\sum_{x \in \mathcal{N}_w} \boldsymbol{h}(x) \right)^{-1} \left(\sum_{x \in \mathcal{N}_w} \boldsymbol{b}(x) \right) \qquad (21)$$

References

[1] D.H. Brainard and W.T. Freeman. Bayesian color constancy. *J. Opt. Soc. Amer.-A*, 14(7):1393–1411, July 1997.

[2] B.A. Buluswar, S.D.; Draper. Color recognition in outdoor images. In *IEEE International Conference on Computer Vision*, January 1998.

[3] J. Cohen. Dependency of the spectral reflectance curves of the munsell color chips. *Psychon. Sci.*, 1:369–370, 1964.

[4] 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.*, 10(10):2148–2165, October 1993.

[5] G. Finlayson, M. Drew, and Funt B. Diagonal transforms suffice for color constancy. In *IEEE International Conference on Computer Vision*, pages 163–171, 1993.

[6] D. Forsyth. A novel approach for color constancy. *International Journal of Computer Vision*, 5:5–36, 1990.

[7] D.A. Forsyth. Sampling, resampling and colour constancy. In *IEEE Computer Vision and Pattern Recognition*, pages I:300–305, 1999.

[8] B.V. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough? In *European Conference on Computer Vision*, 1998.

[9] G. Healey. Segmenting images using normalized color. *IEEE Trans. System, Man and Cybernetics*, 22(1):64–73, January 1992.

[10] T. J. Jansen. *Solar Engineering Technology*. Prentice-Hall, Inc. New Jersey, 1985.

[11] D. B. Judd, D. L. MacAdam, and G. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Amer.*, 54:1031–1040, 1964.

[12] L. T. Maloney and B. Wandell. A computational model of color constancy. *J. Opt. Soc. Amer.*, 1(1):29–33, January 1986.

[13] L.T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Amer.-A*, 3(10):1673–1683, October 1986.

[14] P. Meer. Robust techniques for computer vision (tutorial). In *IEEE Computer Vision and Pattern Recognition*, June 1997.

[15] T. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[16] J.P.S. Parkkinen, J. Hallikainen, and Jaaskelainen T. Characteristic spectra of munsell colors. *J. Opt. Soc. Amer.-A*, 6:318–322, 1989.

[17] C. Rosenberg, M. Hebert, and S. Thrun. Color constancy using KL-divergence. In *IEEE International Conference on Computer Vision*, 2001.

[18] D.A. Slater and G. Healey. What is the spectral dimensionality of illumination functions in outdoor scenes? In *IEEE Comupter Vision and Pattern Recognition*, pages 105–110, 1998.

[19] V.C. Smith and J. Pokorny. Spectral sensitivity of the foveal cone photopigments between 400 and 500nm. *Vision Res*, 15:161–171, 1975.

[20] M. Swain and D. Ballard. Color inidexing. *Int. J. Comput. Vision*, 7(11), November 1991.

[21] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of CCD imaging process. In *IEEE International Conference on Computer Vision*, July 2001.

[22] M. Tsukada and Y. Ohta. An approach to color constancy using multiple images. In *IEEE International Conference on Computer Vision*, pages 385–389, 1990.

[23] B.A. Wandell. The synthesis and analysis of color images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 9(1):2–13, January 1987.