Bayesian Image Color Constancy Using EM and Cached Statistics

Charles Rosenberg

February 24, 2000

Problem Description

The values measured by image capture devices do not directly represent the surface reflectance properties of objects.

The measured pixels values for channel k at location (x, y), denoted as $\rho_k(x, y)$, are the product of:

- the incident illumination $E(\lambda)$
- the surface reflectance properties $S(x, y, \lambda)$
- camera sensor channel spectral sensitivity $C_k(\lambda)$

These are all as a function of the wavelength of the incident light λ, integrated over the visible spectrum, ω:

$$\rho_k(x, y) = \int_\omega E(\lambda)S(x, y, \lambda)C_k(\lambda)d\lambda$$

Color Constancy

The goal of color constancy is to recover the original surface reflectance properties, $S(x, y, \lambda)$, regardless of the incident illumination $E(\lambda)$ to compute a color feature value which is *invariant* of incident illumination.

If $E(\lambda)$ can be estimated, then it is possible to determine the sensor response $\rho_k(x, y)$ under some canonical illuminant and some canonical camera.

Finlayson (1993) demonstrated that with a relatively mild set of assumptions all that needs to be estimated are the three values α, β, and γ in order to recover surface reflectance information:

$$\begin{aligned} R = \rho_r(x, y) = \alpha S_r(x, y) \\ G = \rho_g(x, y) = \beta S_g(x, y) \\ B = \rho_b(x, y) = \gamma S_b(x, y) \end{aligned}$$

Algorithm Outline

Many color constancy, algorithms, Land (1977), Maloney (1986), Gershon (1987), Finlayson (1995), Funt (1996), Finlayson (1998), Funt (1999), are discussed in the literature. All of these algorithms consist of two steps:

- an estimation step
- a correction step

In the estimation step an assumption is made about the distribution of observed pixel color values which is in turn used to estimate the scene illuminant properties.

In the correction step, the illumination estimate is used to calculate a correction factor which is then applied to the image to render it as it would be if captured under a canonical illuminant.

Gray World Algorithm

In the gray world algorithm, the assumption is made that the means of each of the three color channels over the entire image form a fixed ratio under the canonical illuminant.

It is often assumed that the means of the surface reflectances are equal, $\overline{S_r} = \overline{S_g} = \overline{S_b}$, which results in the following identity:

$$\frac{1}{\alpha}\overline{\rho_r} = \frac{1}{\beta}\overline{\rho_g} = \frac{1}{\gamma}\overline{\rho_b}$$

The weakness of this algorithm is that the distribution of colors in the original image may not follow the desired ratio. For example, an image may contain an over abundance of green objects as in a picture of a forest.

White Patch Algorithm

In the white patch retinex algorithm the assumption is made that the peak values in each channel represent the maximum possible reflectance of that component of the illuminant, as might occur from a specular reflection in the image or an object originally colored white.

It is often assumed that the maximum over the channel values are equal, $S_r^{max} = S_g^{max} = S_b^{max}$, which results in the following identity:

$$\frac{1}{\alpha}\rho_r^{max}(x, y) = \frac{1}{\beta}\rho_g^{max}(x, y) = \frac{1}{\gamma}\rho_b^{max}(x, y)$$

A weakness is that it may be hard to identify the maximum value given noise and possible clipping of the channel values. Also, the maximum response from a particular channel may not be characteristic of the response of the illuminant.

Parametric Model

A parametric model, a mixture of Gaussians, of the distribution of color channel responses in an image is constructed and that model is used as the basis for computation in lieu of the raw pixel values.

The hidden indicator variable, z_{ij}, representing which Gaussian generated which pixel, the pixels are indexed by the variable i, ranging from 1 to m where m is the number of pixels in the image, the other variables are defined as follows: the variable j indexes the n Gaussians in the mixture and the variable k indexes the color channels (dimensions), ranging from 1 to 3:

$$Pr(\rho_r(i), \rho_g(i), \rho_b(i) \mid z_{ij}, \mu, \sigma) \;\sim$$

$$(2\pi \textstyle\prod_{k=1}^{3} \sigma_{jk}^2)^{-1/2} \exp\left(-\frac{1}{2} \sum_{k=1}^{3} \frac{(\rho_k(i) - \mu_{jk})^2}{\sigma_{jk}^2} \right)$$

EM Formulation

Expectation step:

$$E[z_{ij}] = \frac{\left(\prod_{k=1}^{3} \sigma_{jk}^2\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{k=1}^{3} \frac{(\rho_k(i)-\mu_{jk})^2}{\sigma_{jk}^2}\right)}{\sum_{c=1}^{n}\left(\prod_{k=1}^{3} \sigma_{ck}^2\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{k=1}^{3} \frac{(\rho_k(i)-\mu_{ck})^2}{\sigma_{ck}^2}\right)}$$

Maximization step:

$$\mu_{jk} \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}]\rho_k(i)}{\sum_{i=1}^{m} E[z_{ij}]}$$

$$\sigma_{jk} \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}]\rho_k^2(i) - \left(\sum_{i=1}^{m} E[z_{ij}]\rho_k(i)\right)^2}{\sum_{i=1}^{m} E[z_{ij}]}$$

Modified Algorithms

The modified gray world algorithm uses the estimated Gaussian means instead of the pixel values. Also, the means of all of the Gaussians are normalize to a constant brightness by choosing a scale factor, $\eta_j = (\mu_{jr} + \mu_{jg} + \mu_{jb})^{-1}$, for each Gaussian such that $\eta_j \mu_{jr} = \eta_j \mu_{jg} = \eta_j \mu_{jb} = 1.0$ for all j. The modified version becomes:

$$\frac{1}{\alpha} \sum_{j=1}^{n} \eta_j \mu_{jr} = \frac{1}{\beta} \sum_{j=1}^{n} \eta_j \mu_{jg} = \frac{1}{\gamma} \sum_{j=1}^{n} \eta_j \mu_{jb}$$

In the modified white patch algorithm we look for the channel maximums in the Gaussian cluster means. The modified version becomes:

$$\frac{1}{\alpha} \mu_{jr}^{max} = \frac{1}{\beta} \mu_{jg}^{max} = \frac{1}{\gamma} \mu_{jb}^{max}$$

Prior work by Gershon (1987) attempts to achieve similar results through the use of image segmentation techniques.

Approximate Methods

The input space is uniformly divided up into cubes. Each cube represents a *bin* and statistics are cached and associated with each bin, based on the pixels which fall into a specific bin.

For each bin, bin_i, seven values are cached: N_i, the number of pixels falling into that bin, $\widehat{\rho}_k(i) = \sum \rho_k(x,y)/N_i$, the mean for each of the color component values, k, for the pixels which fall into that bin, and $\widehat{\rho}_k^2(i) = \sum \rho_k^2(x,y)/N_i$ the mean of the squared color component values for the pixels falling into that bin.

We also experimented with bins in which all of the pixels had equal sums, that is $\rho_r(x,y) + \rho_g(x,y) + \rho_b(x,y)$ is equal to a constant. This creates equi-brightness "slabs", which are separated by planes which are perpendicular to the line $R = G = B$.

Approximate EM Formulation

Expectation step:

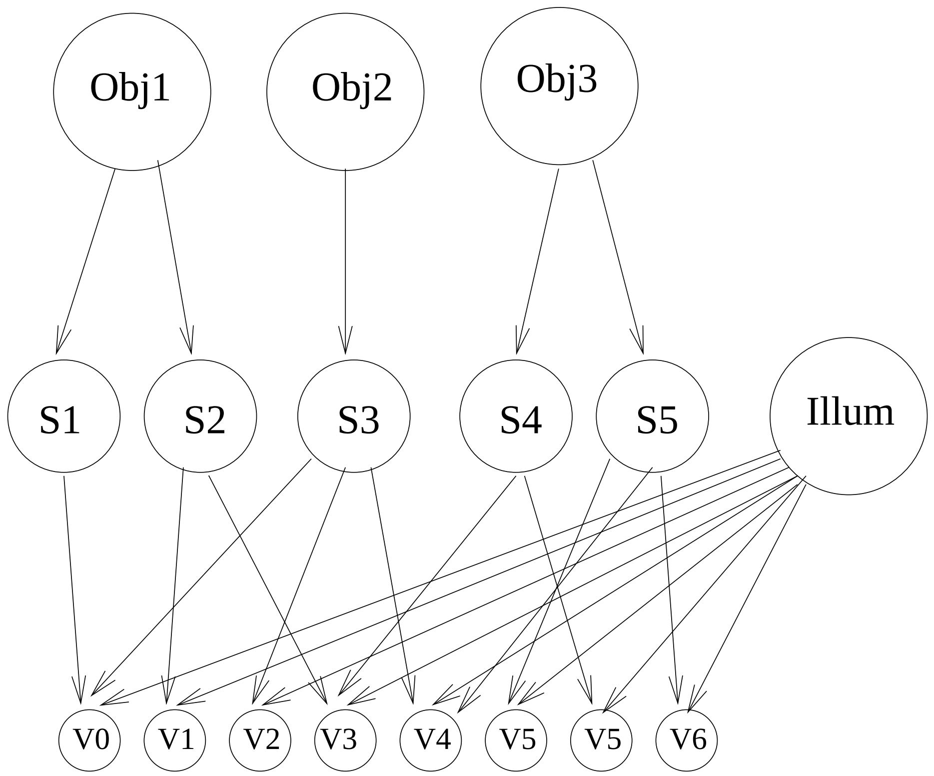
$$E[z_{ij}] = \frac{\left(\prod_{k=1}^{3} \sigma_{jk}^2\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{k=1}^{3} \frac{(\widehat{\rho}_k(i)-\mu_{jk})^2}{\sigma_{jk}^2}\right)}{\sum_{c=1}^{n}\left(\prod_{k=1}^{3} \sigma_{ck}^2\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{k=1}^{3} \frac{(\widehat{\rho}_k(i)-\mu_{ck})^2}{\sigma_{ck}^2}\right)}$$

Maximization step:

$$\mu_{jk} \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}] N_i \widehat{\rho}_k(i)}{\sum_{i=1}^{m} E[z_{ij}] N_i}$$

$$\sigma_{jk} \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}] N_i \widehat{\rho}_k^2(i) - \left(\sum_{i=1}^{m} E[z_{ij}] N_i \widehat{\rho}_k(i)\right)^2}{\sum_{i=1}^{m} E[z_{ij}] N_i}$$

Conditional Dependence Structure



Probabilistic Framework

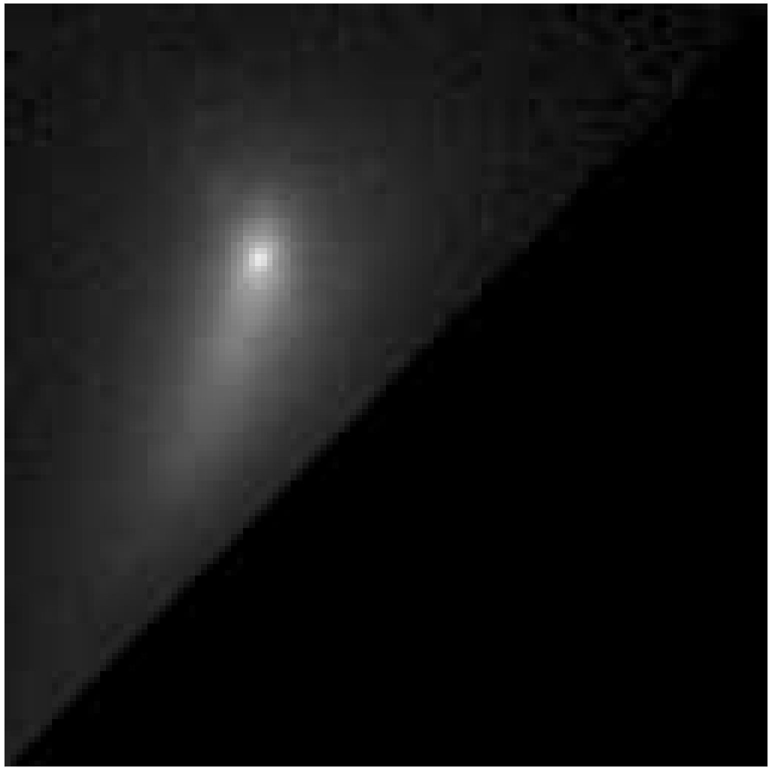
$$P(I_j|\underline{V})\alpha \prod_{i=1}^{n} P(V_i|I_j)P(I_j)$$

$$V_i = S_i \times I_j$$

$$P(V_i|I_j) = P(S_i \times I_j|I_j) = P(S_i|I_j)$$

$$P(I_j|\underline{V})\alpha \prod_{i=1}^{n} P(S_i|I_j)P(I_j)$$

Surface Color Prior



Results and Evaluation Metrics

The error in the illuminant estimated was measured as the Euclidean distance between the chromaticity of the “true” illuminant to the estimated illuminant. The “true” illuminant was estimated from a calibration target in the image.

The chromaticity measure used in this work is a two dimensional quantity calculated from the triplet α, β, γ of illuminant scaling values as follows:

$$C_1 = \frac{\alpha}{\alpha+\beta+\gamma}, C_2 = \frac{\beta}{\alpha+\beta+\gamma}$$

Only two values are needed because $C_3 = 1 - C_1 - C_2$.

Algorithms were also timed to measure relative computational performance.

Data Set

The data set was collected by the Computational Vision Lab at Simon Fraser University:

> Sony DXC-930 3-CCD video camera, gamma correction turned off, data post processed to remove offsets

Images of 11 different household objects on black backgrounds:

> two different colored balls, a book, a disposable coffee cup, a cereal box, flowers, a bottle of bleach, a box of macaroni, some colored rope, a shampoo bottle, a detergent box

Data Set

The images were captured under the following five illuminants:

> Sylvania Halogen, Sylvania Cool White Fluorescent, a Phillips Ultralume Fluorescent, the Macbeth Judge II 5000 Fluorescent, and the Macbeth Judge II 5000 Fluorescent together with a Roscolux 3202 blue filter.

In addition to the selected object, each image also included the Gretag / Macbeth ColorChecker color rendition chart.

Image intensity values were scaled by a factor of 2.5 and values over 255 were clipped to more closely match typical exposure settings.

Unmodified Image | Simple White Patch | EM White Patch | Optimal Estimate

Visual Evaluation

These color images provide an example of algorithm performance for visual inspection. Images in the first row were captured under halogen illumination. Images in the second row were captured under Macbeth 5000 illumination. The columns represent, respectively, from left to right: no correction, simple white patch correction, EM white patch correction with 512 uniform bins, optimal correction.

Experimental Results

Algorithm	Average Distance Error	Standard Deviation	Significance
White Patch Retinex	0.0999	0.0616	—
White Patch EM / Raw Pixels	0.0864	0.0537	0.042
WP EM / 4096 Uniform Bins	0.0847	0.0528	0.025
WP EM / 512 Uniform Bins	0.0826	0.0500	0.012
WP EM / 1024 Projection Bins	0.0859	0.0523	0.035

Distance results for the white patch algorithms, larger distances represent larger errors. The last column labeled "Significance" is the probability that the specific algorithm actually performs worse than the basic algorithm, even though the mean error appears to be less. This probability was estimated using a student t distribution.

Experimental Results

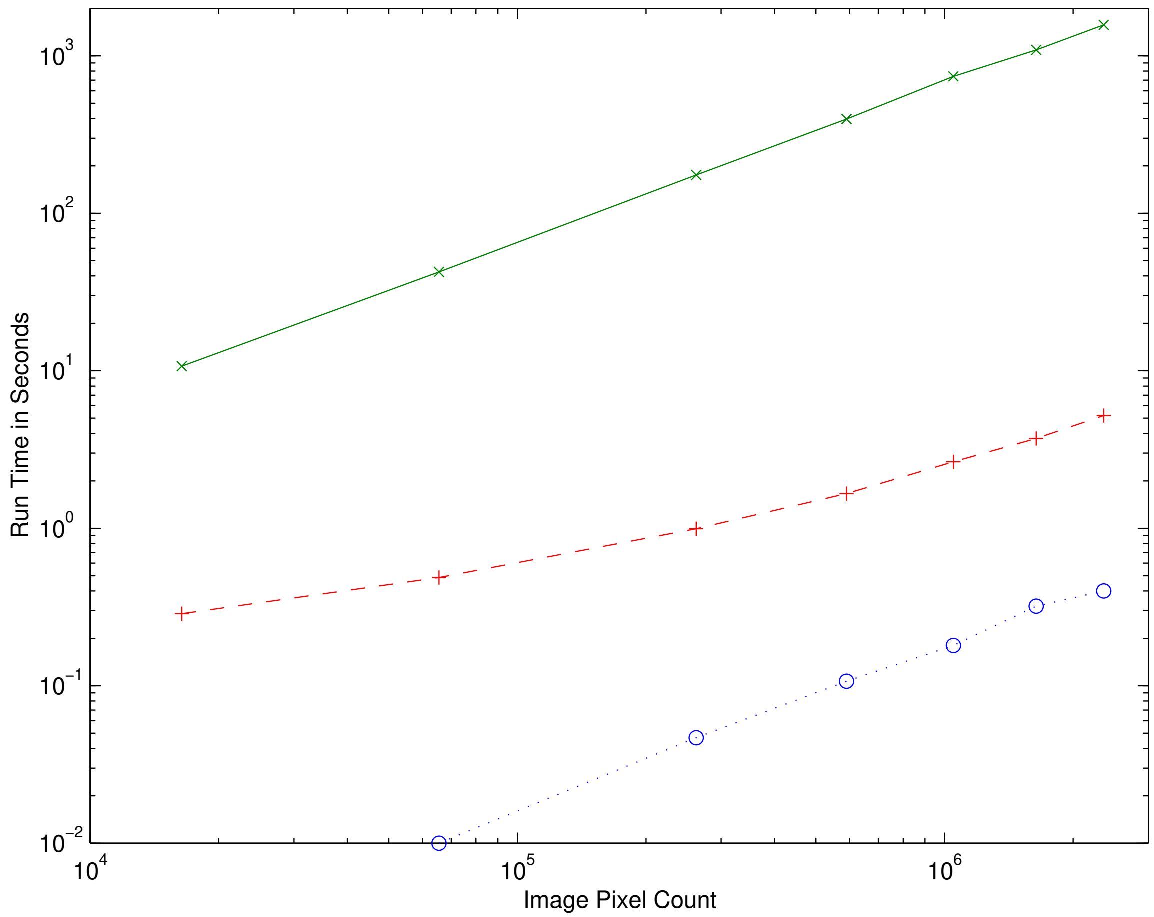
Algorithm	Average Distance Error	Standard Deviation	Significance
Gray World	0.0833	0.0568	—
Gray World EM / Raw Pixels	0.0808	0.0496	0.37
GW EM / 4096 Uniform Bins	0.1156	0.0655	0.00
GW EM / 512 Uniform Bins	0.1123	0.0579	0.00
GW EM / 1024 Projection Bins	0.0773	0.0496	0.20

Distance results for the gray world algorithms.

Experimental Results

Image Size	White Patch	White Patch EM 4096 Bins	White Patch EM Raw Pixels
128 x 128	0.00	0.29	10.70
256 x 256	0.01	0.49	42.49
512 x 512	0.06	0.99	175.40
768 x 768	0.11	1.66	396.68
1024 x 1024	0.18	2.65	740.66
1280 x 1280	0.32	3.72	1089.98
1536 x 1536	0.40	5.20	1575.80

For a of 512 x 512 image, the binned EM version is 16.5 times slower than the basic algorithm, but still takes less than one second for a typical image. The full EM version is 2900 times slower than the basic version, and takes nearly 3 minutes to process a typical size image.



The dotted line plots data from the basic version of the algorithm.
The dashed line plots data from the EM version of the algorithm with 4096 bins.
The solid line plots data from the version of the algorithm utilizing raw pixels.

Related Work

Finlayson (1995) utilizes constraints about the distribution of the extreme colors in an image. The constraints are formed by knowledge of typical object surfaces and illuminants in the world. A search procedure is used to determine the transformation which best satisfies the given constraints.

Funt (1996) uses a neural network which takes as input the image color histogram and outputs estimated chromaticities. Recent work by Funt (1999) uses bootstrapping to eliminate the need for expensive training data.

The “retinex” family of algorithms was originated by Land (1977). The lightest regions in the image provide a measurement of the parameters of the original scene illumination.

Gershon (1987) uses image segmentation techniques are used to divide an image into regions.

Conclusions and Future Work

- Constructing and using a parametric model of the distribution of image color pixel values can be utilized by color constancy algorithms to improve their performance.

- Next line of investigation in this work will be to experiment with mixture models which use the full covariance matrix. Our performance results suggest that performance could be improved in that manner.

- Also interesting is the projected binning algorithm which seemed to improve both algorithm accuracy as well as computational performance. This suggests that for certain applications a carefully chosen set of bins might provide some of the benefit of using the full covariance matrix without the complexity.

References

Finlayson, G. & Drew, M. S. & Funt, B. (1993). Diagonal Transforms Suffice for Color Constancy. *IEEE Proceedings: International Conference on Computer Vision*, 164-171, Berlin.

Finlayson, G. & Funt, B. & Barnard, J. (1995) Color Constancy Under a Varying Illumination. *Proceedings of the Fifth Intl. Conf. on Computer Vision.*

Finlayson, G. & Schiele, B. & Crowley, J. (1998) Comprehensive Color Image Normalization. *Fifth European Conference on Computer Vision*. Freiburg, Germany,

Funt, B. & Cardei, V. & Barnard, K. (1996) Learning Color Constancy. *Proc. IS&T/SID Fourth Color Imaging Conf.*, 58-60. Scottsdale.

Funt, B. & Barnard, K. & Martin, L. (1998) Is colour constancy good enough? *5th European Conference on Computer Vision*, 445-459.

Funt, B. & Cardei, V. C. (1999) Bootstrapping Color Constancy. *Proc. SPIE: Electronic Imaging IV*, 3644.

Land, E.H. (December 1977) The retinex theory of color vision. *Scientific American*, 108-128.

Maloney, L. T. and Wandell, B. A. (1986) Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc America A*, 3:29-33.

Moore, A. W. (1999) Very Fast EM-based Mixture Model Clustering using Multiresolution kd-trees. *Advances in Neural Information Processing Systems 11*.