Image Color Constancy
Using EM and Cached Statistics

Charles Rosenberg

Department of Computer Science

Carnegie Mellon University

ICML 2000

Presentation Outline

- Problem Description
- Basic Algorithms
- Generative Model
- Modified Algorithms
- Approximate Methods
- Experimental Results
- Conclusions and Future Work

Problem Description

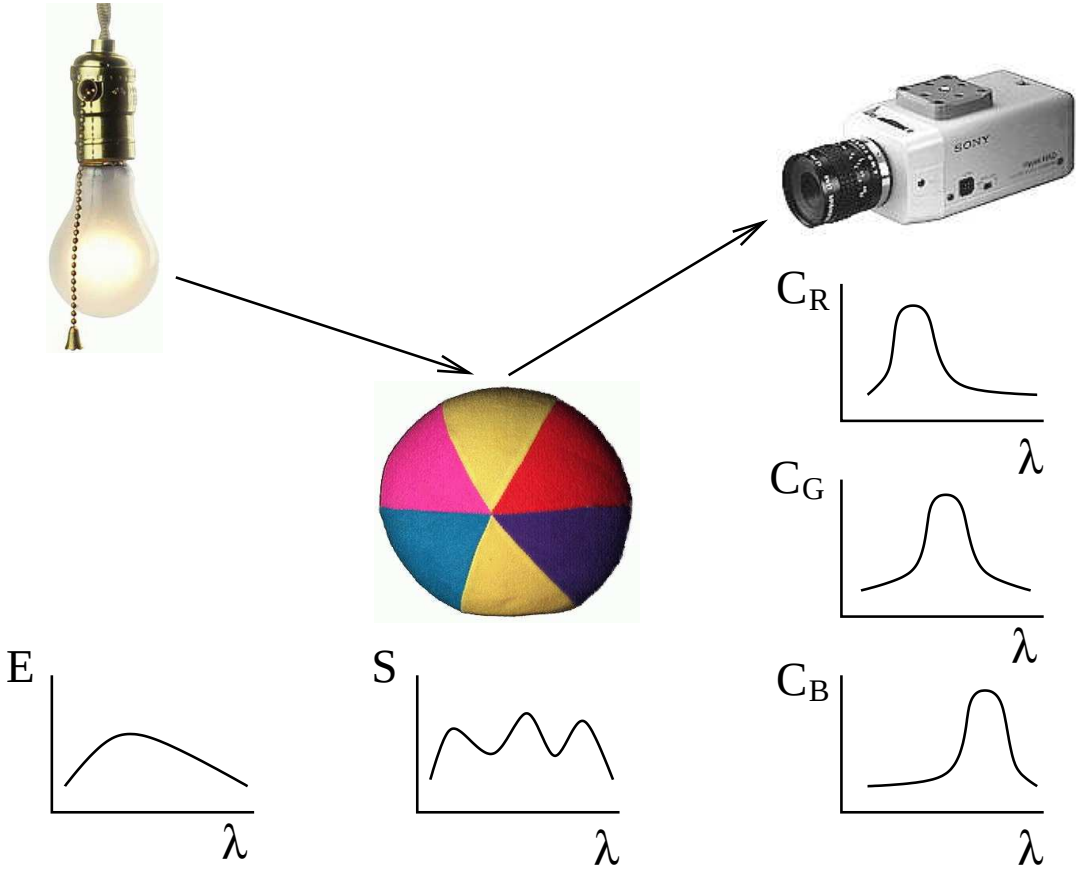


Image capture devices do not directly measure the surface reflectance properties of objects.

Problem Description

The measured pixels values are a function of:

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

The pixel value for channel k at location (x, y), denoted as $\rho_k(x, y)$, is:

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

Example Raw Images

Halogen

Fluorescent & Blue Filter

MB5000 Fluorescent

Ultralume Fluorescent

Cool White Fluorescent

Color Constancy

The goal of color constancy is to recover a color feature value which is *invariant* of incident illumination, preferably $S(x, y, \lambda)$.

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

Finlayson (1993) demonstrated that, with a mild set of assumptions, a simple model using only three values, α, β, and γ, can be utilized:

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

Gray World Algorithm

Here the assumption is made that the means of each of the three color channels over the entire image form a fixed ratio:

$$\frac{1}{\alpha}\overline{\rho_r} = \overline{S_r}, \frac{1}{\beta}\overline{\rho_g} = \overline{S_g}, \frac{1}{\gamma}\overline{\rho_b} = \overline{S_b}$$

The weakness of this algorithm is that the distribution of colors in the original image may not follow the desired ratio.

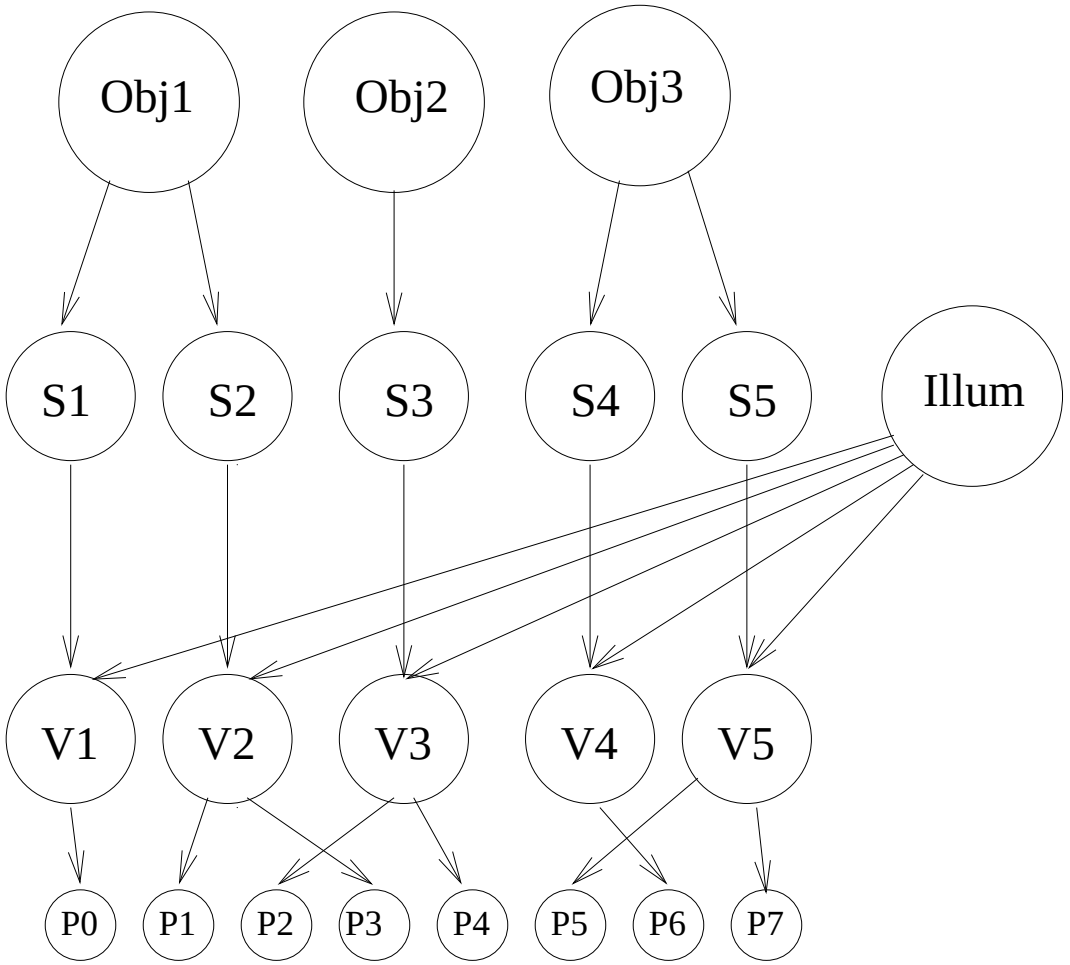
White Patch Algorithm

Here the assumption is made that the peak values in each channel represent the maximum possible illuminant reflectance component and form a fixed ratio over the channel values:

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

A weakness is that it may be hard to identify the maximum value given noise and possible clipping of the channel values.

Generative Model Structure



Parametric Model

A 3D mixture of Gaussians model, of the distribution of color channel responses in an image is constructed :

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

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

z_{ij}	hidden indicator variable, representing which Gaussian generated which pixel
i	image pixel index, ranging from 1 to m
j	Gaussian mixture index, ranging from 1 to n
k	color channel index (dimension), ranging from 1 to 3

EM Formulation

Expectation step:

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

Maximization step:

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

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

Modified Algorithms

The modified gray world algorithm uses the estimated Gaussian means instead of the pixel values. The modified version becomes:

$$\frac{1}{\alpha \cdot n} \sum_{j=1}^{n} \mu_{jr} = \overline{S_r}, \frac{1}{\beta \cdot n} \sum_{j=1}^{n} \mu_{jg} = \overline{S_g}, \frac{1}{\gamma \cdot n} \sum_{j=1}^{n} \mu_{jb} = \overline{S_b}$$

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

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

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
$\widehat{\rho}_k(i)$	$\widehat{\rho}_k(i) = \sum \rho_k(x, y)/N_i$, the mean for each of the color component values
$\widehat{\rho}_k^2(i)$	$\widehat{\rho}_k^2(i) = \sum \rho_k^2(x, y)/N_i$, the mean of the squared color component values for the pixels falling into that bin

Approximate EM Formulation

Expectation step:

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

Maximization step:

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

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

Data Set

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

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

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, the Macbeth Judge II 5000 Fluorescent together with a Roscolux 3202 blue filter.

Example Corrected Images - WP EM

err = 0.0505 err = 0.0490 err =0.0160 err = 0.0200 err = 0.0301

Example Corrected Images - WP EM

err = 0.0505 err = 0.0490 err =0.0160 err = 0.0200 err = 0.0301

Experimental Results for White Patch Algorithms

Algorithm	Avg. Distance Error	Std. Deviation	Significance
White Patch Retinex	0.0992	0.0611	—
White Patch EM / Raw Pixels	0.0665	0.0469	4.99×10^{-7}
WP EM / 4096 Uniform Bins	0.0644	0.0445	1.83×10^{-7}
WP EM / 512 Uniform Bins	0.0639	0.0436	1.44×10^{-7}
WP EM / 1024 Projection Bins	0.0637	0.0475	1.94×10^{-7}
Simple Mean	0.1492	0.1218	2.57×10^{-6}

The chromaticity measure used in this work is calculated as follows:

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

The error in the illuminant estimated was measured as follows:

$$Error = \sqrt{(C_1^{est} - C_1^{true})^2 + (C_2^{est} - C_2^{true})^2}$$

Experimental Results for Gray World Algorithms

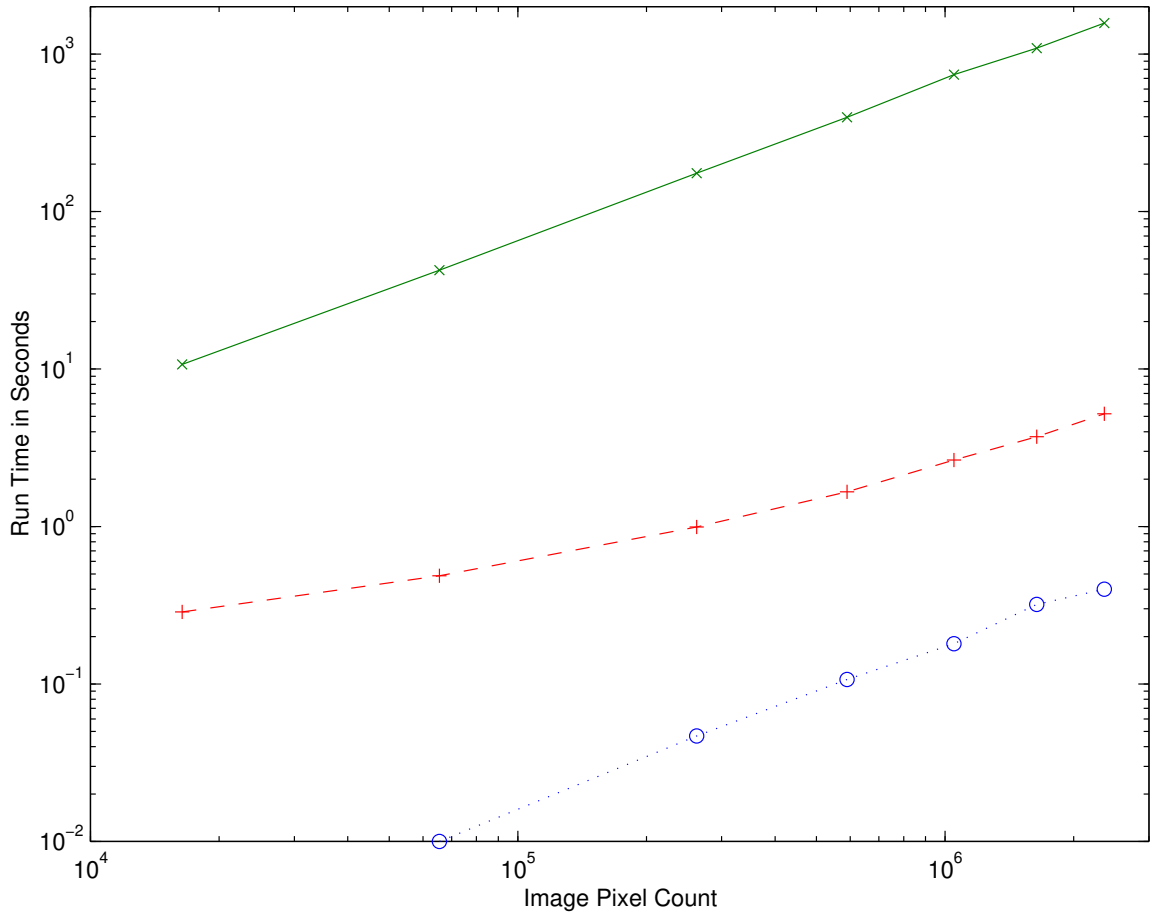
Algorithm	Avg. Distance Error	Std. Deviation	Significance
Gray World	0.0625	0.0420	—
Gray World EM / Raw Pixels	0.0611	0.0400	0.169
GW EM / 4096 Uniform Bins	0.0595	0.0380	0.053
GW EM / 512 Uniform Bins	0.0594	0.0391	0.050
GW EM / 1024 Projection Bins	0.0620	0.0444	0.322
Simple Mean	0.1492	0.1218	6.72×10^{-10}

Timing 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 512 x 512 image:

- the binned EM version is 16.5 times slower than the basic algorithm, but still takes less than one second
- the full EM version is 2900 times slower than the basic algorithm and takes nearly 3 minutes



- 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 by knowledge of typical object surfaces and illuminants in the world. A search procedure is used to determine parameters which best satisfy 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 to divide an image into regions.

Conclusions

- 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.
- Using cached statistics to build a mixture models greatly reduces computation complexity and can yeild high quality results.

Future Work

- Next line of investigation will be to experiment with mixture models which use the full covariance matrix.
- Future work will also include using a Bayesian framework to estimate the illumination models.

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*.