Learning Outdoor Color Classification from Just One Training Image

Roberto Manduchi

University of California, Santa Cruz

Abstract. We present an algorithm for color classification with explicit illuminant estimation and compensation. A Gaussian classifier is trained with color samples from just one training image. Then, using a simple diagonal illumination model, the illuminants in a new scene that contains some of the same surface classes are estimated in a Maximum Likelihood framework using the Expectation Maximization algorithm. We also show how to impose priors on the illuminants, effectively computing a Maximum-A-Posteriori estimation. Experimental results show the excellent performances of our classification algorithm for outdoor images. [1]

1 Introduction

Recognition (or, more generally, classification) is a fundamental task in computer vision. Differently from clustering/segmentation, the classification process relies on prior information, in the form of physical modeling and/or of training data, to assign labels to images or image areas. This paper is concerned with the classification of outdoor scenes based on color. Color features are normally used in such different domains as robotics, image database indexing, remote sensing, tracking, and biometrics. Color vectors are generated directly by the sensor for each pixel, as opposed to other features, such as texture or optical flow, which require possibly complex pre–processing. In addition, color information can be exploited at the local level, enabling simple classifiers that do not need to worry too much about contextual spatial information.

Color–based classification relies on the fact that a surface type is often uniquely characterized by its reflectance spectrum: different surface types usually have rather different reflectance characteristics. Unfortunately, a camera does not take direct reflectance measurements. Even neglecting specular and non–Lambertian components, the spectral distribution of the radiance from a surface is a function of the illuminant spectrum (or spectra) as much as of the surface reflectance. The illuminant spectrum is in this context a nuisance parameter, inducing an undesired degree of randomness to the perceived color of a surface. Unless one is interested solely in surfaces with a highly distinctive reflectance (such as the bright targets often used in laboratory robotics experiments), ambiguity and therefore misclassification will arise when surfaces are illuminated by “unfamiliar” light.

[1] This work was supported by DARPA through subcontract 1235249 from JPL.

One way to reduce the dependence on the illuminant is to use more training data and sophisticated statistical models to represent the color variability. This is feasible if the space of possible illuminants is not too broad, meaning that we can hope to sample it adequately. For example, in the case of outdoor scenes (which are of interest to this work), the spectrum of the illuminant (direct sunlight or diffuse light and shade) can be well modeled by a low–dimensional linear space. Thus, by collecting many image samples of the surfaces of interest under all expected light conditions, one may derive the complete statistical distribution of colors within each class considered, and therefore build a Bayesian classifier that can effectively cope with variation of illumination. This approach was taken by the author and colleagues for the design and implementation of the color–based terrain typing subsystem of the eXperimental Unmanned Vehicle (XUV) DEMO III [5], which provided excellent classification performances. Unfortunately, collecting and hand–labeling extensive training data sets may be difficult, time–consuming, and impractical or impossible in many real–world scenarios. This prompted us to study an orthogonal approach, relying on a model–based, rather than exemplar–based, description of the data. Our algorithm aims to decouple the contribution of the reflectance and of the illumination components to the color distribution within each surface class, and to explicitly recover and compensate for variations of the illuminant (or illuminants) in the scene. Both components (reflectance and illuminant) are modeled by suitable (and simple) statistical distributions. The contribution of reflectance to the color distribution is learned by observing each class under just one "canonical" illuminant, possibly within a single training image. To model the contribution of illumination, one may either directly hard–code existing chromaticity daylight curves [4] into the system, or learn the relevant parameters from a data set of observations of a fixed target (such as a color chart) under a wide variety of illumination conditions. Note that illuminant priors learning is performed once and for all, even before choosing the classes of interest.

The estimation of the illuminants present in the scene, together with the determination of which illuminant impinges on each surface element, is performed by a Maximum–A–Posteriori (MAP) algorithm based on the distributions estimated in the training phase. Our formulation of the MAP criterion is very similar to the one by Tsin et al. [3]. Our work, however, differs from [3] in two main aspects. Firstly, [3] requires that a number of images of the same scene, containing the surface types of interest, are collected by a fixed camera under variable lighting conditions. While this training procedure may be feasible for surveillance systems with still cameras, it is impractical for other applications (such as robotics). As mentioned earlier, our system only requires one image containing the surfaces of interest under a single illuminant. Secondly, our algorithm is conceptually and computationally simpler than [3]. Instead of an ad–hoc procedure, we rely on the well–understood Expectation Maximization algorithm for illuminant parameter estimation. A simple modification of the EM algorithm allows us to include the prior distribution of the illuminant parameters for a truly Bayesian estimation. Illuminant priors are critical when dealing with scenes containing surfaces that were not used during training. Without prior knowledge of the actual statistics of illuminant parameters, the system would be tempted to "explain too much" of the scene, that is, to model the color of a never seen before surface as the transformation of a known reflectance under a very unlikely light. To further shield the algorithm from the influence of "outlier" surfaces, we also augment the set of classes with a non–informative class distribution, a standard procedure in similar cases. The price to pay for the simplicity of our algorithm is a less accurate model of color production than in [3], which potentially may lead to lower accuracy in the illuminant compensation process. We use the diagonal model [8] to relate the variation of the illuminant spectrum to the perceived color. It is well known that a single diagonal color transformation cannot, in general, accurately predict the new colors of different surface types. However, we argue that the computational advantages of using such a simple model largely offset the possibly inaccurate color prediction. Note that other researchers have used the diagonal color transformation for classification purposes (e.g. [12]).

2 The Algorithm

Assume that K surface classes of interest have been identified, and that training has been performed over one or more images, where all samples used for training are illuminated by the same illuminant. Let $p(c|k)$ denote the conditional likelihood over colors c for the class model k, as estimated from the training data. The total likelihood of color c is thus

$$p(c) = \sum_{k=1}^{K} P_K(k) p(c|k) \tag{1}$$

where $P_K(k)$ is the prior probability of surface class k. In general, a scene to be classified contains a number of surfaces, some (but not all) of which belong to the set of classes used for training, and are illuminated by one or more illuminants which may be different from the illuminant used for training. Assume there are L possible illuminant types in the scene. Let c be the color of the pixel which is the projection of a certain surface patch under illuminant type l. We will denote by $F_l(c)$ the operator that transforms c into the color that would be seen if the illuminant of type l had the same spectrum as the one used for training, all other conditions being the same (remember that only one illuminant is used for training). Then, one may compute the conditional likelihood of a color c in a test image given surface class k, illuminant type l, and transformation F_l:

$$p_F(c|k,l) = p(F_l(c)|k)|J(F_l)|_c \tag{2}$$

where $|J(F_l)|_c$ is the absolute value of the Jacobian of F_l at c.

We will begin our analysis by making the following assumptions: 1) The surface class and illuminant type at any given pixel are mutually independent random variables; 2) The surface class and illuminant type at any given pixel are independent of the surface classes and illuminant types at nearby pixels; 3)

The color of any given pixel is indepedent of the color of nearby pixels, even when they correspond to the same surface class and illuminant type; 4) Each surface element is illuminated by just one illuminant. Assumption 1) is fairly well justified: the fact that a surface is under direct sunlight or in the shade should be independent of the surface type. It should be noticed, however, that in case of "rough" surfaces (e.g., foliage), self–shading will always be present even when the surface is under direct sunlight. Assumption 2) is not very realistic: nearby pixels are very likely to belong to the same surface class and illuminant type. This is indeed a general problem in computer vision, by no means specific to this particular application. We can therefore resort to standard approaches to deal with spatial coherence [16, 15]. The "independent noise" assumption 3) is perhaps not fully realistic (nearby pixels of the same smooth surface under the same illuminant will have similar color), but, in our experience, it is a rather harmless, and computationally quite convenient, hypothesis. Assumption 4) is a very good approximation for outdoor scenes, where the only two illuminants (excluding inter–reflections) are direct sunlight and diffuse light (shade) [1].

With such assumptions in place, we may write the total log–likelihood of the collection of color points in the image, C, given the set of L transformations F_l, as

$$\begin{aligned} L_F(C) &= \sum_x \log \sum_{l=1}^{L} \sum_{k=1}^{K} P_{KL}(k,l) p_F(c(x)|k,l) \qquad (3) \\ &= \sum_x \log \sum_{l=1}^{L} \sum_{k=1}^{K} P_K(k) P_L(l) p(F_l(c(x))|k) |J(F_l)|_{c(x)} \end{aligned}$$

where $c(x)$ is the color at pixel x, and we factorized the joint prior distribution of surface class and illuminant type ($P_{KL}(k,l) = P_K(k)P_L(l)$) according to Assumption 1. Note that the first summation extends over all image pixels, and that L and K are the number of possible illuminants and surface classes, which are supposed to be known in advance. In our experiments with outdoor scenes, we always assumed that only two illuminants (sunlight and diffuse light) were be present, hence L=2.

Our goal here is to estimate the L transformations $\{F_l\}$ from the image C, knowing the conditional likelihoods $p(c|k)$ and the priors $P_K(k)$ and $P_L(l)$. Once such transformations have been estimated, we may assign each pixel x a surface class k and illuminant type l as by

$$\{k,l\} = \arg\max P_K(k) P_L(l) p(F_l(c(x))|k) |J(F_l)|_{c(x)} \qquad (4)$$

We will first present a ML strategy to determine $\{F_l\}$, which maximizes the total image log–likelihood (3). In Section 2.3, we will show how existing priors on the color transformations can be used in a MAP setting.

2.1 Estimating the Illuminant Parameters

As mentioned in the Introduction, we will restrict our attention to diagonal transformations of the type $F_l(c) = D_l c$, where $D_l = \text{diag}(d_{l,1}, d_{l,2}, d_{l,3})$. Note that in this case, $|J(F_l)|_{c(x)} = |d_{l,1}d_{l,2}d_{l,3}|$. To make the optimization problem more tractable, we will assume that $p(c|k) \in \mathcal{N}(\mu_k, \Sigma_k)$. While this Gaussian assumption may not be acceptable in general (especially for multimodal color distributions), it has been shown in the literature that mixtures of Gaussians can successfully model color distributions [5, 19]. The extension of our optimization algorithm to the case of Gaussian mixtures is trivial.

The optimal set of $3L$ color transformation coefficients $\{d_{l,m}\}$ can be found using Expectation Maximization (EM) [13]. EM is an iterative algorithm that re–estimates the model parameters in such a way that the total image log–likelihood, $L_F(C)$, is increased at each iteration. It is shown in Appendix A that each iteration is comprised of L independent estimations, one per illuminant type. For the l–th illuminant type, one needs to compute

$$\{d_{l,m}\} = \arg\max \left[u_l \sum_{\bar{m}=1}^{3} \log|d_{l,\bar{m}}| - 0.5\ d_l' G_l d_l + H_l' d_l \right] \tag{5}$$

where $d_l = (d_{l,1}, d_{l,2}, d_{l,3})'$. The scalar u_l, the 3x3 matrix G_l and the 3x1 vector H_l are defined in Appendix A, and are re–computed at each iteration. Our task now is to minimize (5) over $(d_{l,1}, d_{l,2}, d_{l,3})$. Note that the partial derivatives with respect to $d_{l,m}$ of the function to be maximized can be computed explicitely. Setting such partial derivatives to zero yields the following system of quadratic equations for $m = 1, 2, 3$:

$$\sum_s G_{l,m,s} d_{l,s} d_{l,m} - H_{l,m} d_{l,m} - u_l = 0 \tag{6}$$

While this system cannot be solved in closed form, we note that if two variables (say, $d_{l,1}$ and $d_{l,2}$) are kept fixed, then the partial derivative of (5) with respect to the third variable can be set to 0 by solving a simple quadratic equation. Hence, we can minimize (5) by using the Direction Set method [14], i.e. iterating function minimization over the three axes until some convergence criterion is reached. Note that this is a very fast maximization procedure, and that its complexity is independent of the number of pixels in the image.

2.2 Outliers

Any classification algorithm should account for "unexpected" situations that were not considered during training, by recognizing outlier or "none of the above" points. A popular strategy treats outliers as an additional class, for which a prior probability and a least informative conditional likelihood are defined [20]. For example, if one knows that the observables can only occupy a bounded region in the measurement space (a realistic assumption in the case of color features), one may allocate a uniform outlier distribution over such a region.

In our work, we defined an outlier surface class with an associated uniform conditional likelihood over the color cube $[0:255]^3$. Note that the outlier class contributes to the parameter estimation only indirectly, in the sense that equation (2) does not apply to it. In other words, color transformations do not change the conditional likelihood given the outlier class.

2.3 Imposing illuminant priors

A problem with the approach detailed above is that the algorithm, when presented with scenes that contain surfaces with very different colors from the ones used for training, may arbitrarily "invent" unusual illuminant in order to maximize the scene likelihood. Illuminant spectral distributions in outdoor scene are rather constrained [4, 1]; this observation should be exploited to reduce the risk of such catastrophic situations. Prior distributions on the parameters can indeed be plugged into the same EM machinery used for Maximum Likelihood by suitably modifying the function to be maximized at each iteration. More precisely, as discussed in [13], imposing a prior distribution on the parameter d translated into adding the term $\log p_D(\{D_l\})$ to the function $Q\left(\{D_l\}, \{D_l^0\}\right)$ in Appendix A before the maximization step. Assuming that the different illuminants in the scene are statistically independent, we may write

$$\log p_D(\{D_l\}) = \sum_{l}^{L} \log p_d(d_l) \tag{7}$$

where $d_l = (d_{l,1}, d_{l,2}, d_{l,3})'$. We will assume that all L illuminant types have the same prior probability.

One way to represent the illuminant priors could be to start from the CIE parametric curve, thus deriving a statistical model for the matrices D_l. Another approach, which we used in this work, is to take a number of pictures of the same target under a large number of illumination conditions, and analyze the variability of the color transformation matrices. For example, in our experiments we took 39 pictures of the Macbeth color chart at random times during daylight over the course of one week. For each corresponding color square in each pair of images in our set, we computed the ratios of the r, g and b color components in the two images. The hope is that the ensemble of all such triplets can adequately model the distribution of diagonal color transformations. We built a Gaussian model for the prior distribution of d by computing the mean μ_d and covariance Σ_d of the collected color ratio triplets. These Gaussian priors can be injected into our algorithm by modifying (5) into

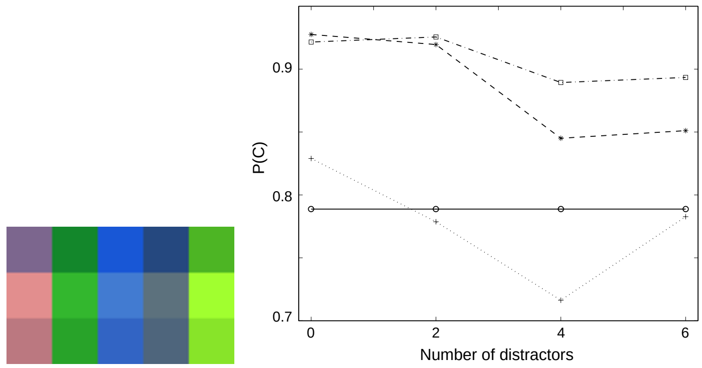
$$\{d_{l,m}\} = \arg\max \left[u_l \sum_{\bar{m}=1}^{3} \log |d_{l,\bar{m}}| - 0.5\ d_l'(G_l + \Sigma_d^{-1})d_l + (H_l' + \mu_d' \Sigma_d^{-1})d_l \right] \tag{8}$$

Fortunately, even with this new formulation of the optimization criterion, we can still use the Direction Set algorithm for maximization, as in Section 2.1.

3 Experiments

3.1 Macbeth Color Chart Experiments

In order to provide a quantitative evaluation of our algorithm's performance, we first experimented using mosaics formed by color squares extracted from



(a) (b)

Fig. 1. (a) The color squares used for training (center row) and for testing(top and bottom row) in the Macbeth color chart experiment. (b) The probability of correct match $P(C)$ as a function of the total number of distractors for the Macbeth color chart experiment without illuminant compensation (solid line), with ML illuminant compensation (dotted line), with MAP illuminant compensation without the outlier class (dashed line), and with MAP illuminant compensation using the outlier class (dashed–dotted line).

pictures of the GretagMacbeth™ chart under different illuminations, taken by a Sony DSC–S75 camera[2]. We picked 5 colors from the chart (No. 2, 14, 3, 22,11) as representative of 5 classes of interest. In Figure 1(a) we show the five colors as seen under evening light (after sunset, top row), direct afternoon sunlight (center row), and diffuse (shade) afternoon light (bottom row). We ran a number of experiment by training the system over the color squares in the middle row of Figure 1(a), and testing on a mosaic composed by color squares in the the top and bottom row, as well as by other colors in the chart ("distractors"). More precisely, for each test we formed two samples, one from the top row and one from the bottom row of Figure 1(a). Each sample had a number (randomly chosen between 0 and 5) of color squares, randomly chosen from those in the corresponding row, making sure that at least one of such two samples was not empty. Then, we augmented each sample with a number of randomly selected distractors, that were not present in the training set. The test image is the union of the two samples. We ran 100 tests for each choice of the number of distractors per sample (which varied from 0 to 3). At each test, we first tried to assign each non–distractor color in the test image to one of the colors in the training set using Euclidean distance in color space. The ratio between the cumulative number of correct matches and the cumulative number of non–distractors squares in the test images over all 100 tests provides an indication of the probability of correct match without illuminant compensation. Such value is shown in Figure 1(b) by

[2] In order to roughly compensate for the camera's gamma correction, we squared each color component before processing.

solid line. Obviously, these results do not depend on the number of distractors. In order to test our illumination compensation algorithm, we added some "virtual noise" to the colors within each square, by imposing a diagonal covariance matrix with marginal variances equal to 10^8 (the reader is reminded that the color values were squared to reduce the effect of camera gamma correction). This artifice is necessary in this case because the color distribution in the squares had extremely small variance, which would create numerical problems in the implementation of the EM iterations. We didn't need to add noise in the real–world tests of Section 3.2. The number L of illuminants in the algorithm was set to 2. The probability of correct match after illuminant compensation using the ML algorithm of Section 2.1 (without exploiting illuminant priors and without the outlier class), of the MAP algorithm of Section 2.3 (without the outlier class), and of the MAP algorithm using the outlier class, are shown in Figure 1(b) by dotted line, dashed line, and dashed–dotted line respectively. Note that the distractors contribute to the determination of the illuminant parameters, and therefore affect the performance of the illuminant compensation system, as seen in Figure 1(b). Distractors have a dramatic negative effect if the illuminant priors are not taken into consideration, since the system will "invent" unlikely illumination parameters. However, the MAP algorithm is much less sensitive to distractors; if, in addition, the outlier class is used in the optimization, we see that illuminant compensation allows one to increase the correct match rate by 5–13%.

3.2 Experiments with Real–World Scenes

We tested our algorithm on a number of outdoor scenes, with consistently good results. We present two experiments of illuminant–compensated color classification in Figure 2 and 3. Figure 2 (a) and (b) shows two images of the same scene under very different illuminants. Three color classes were trained over the rectangular areas shown in the image of Figure 2 (a), with one Gaussian mode per class. The result of classification after illuminant compensation are shown in Figure 2 (c). Pixels colored in blue are considered outliers. Note that for this set of images, we extended our definition of outlier to overexposed pixels (i.e., pixels that have one or more color components equal to 255), which typically correspond to the visible portion of the sky. Figure 2 (d) shows the illuminant–compensated image. It is seen that normalization yields colors that are very similar to those in the training image. The assignment of illuminant types is shown in Figure 2 (f). Figure 2 (e) shows the result of classification without illuminant compensation. In this case, large image areas have been assigned to the outlier class, while other areas have been misclassified. Comparing these results with those of Figure 2 (c) shows the performance improvement enabled by our illuminant compensation algorithm.

Our second experiment is described in Figure 3. Figure 3 (a) shows the training image (three classes were trained using the pixels within the marked rectangles), while Figure 3 (b) shows the test image. The results of classification after illuminant estimation are shown in Figure 3 (c). Pixels colored in blue are

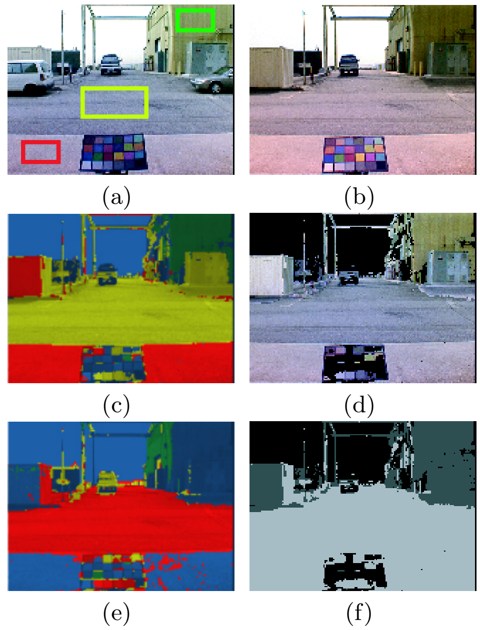
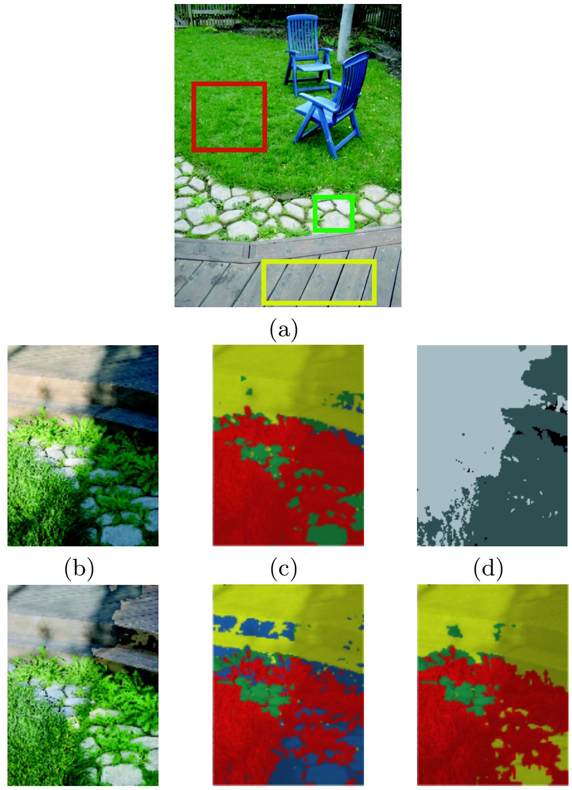


Fig. 2. Color classification experiments: (a): training image; (b) test image; (c) classification with illuminant compensation (blue indicates outliers); (d) illuminant–compensated version of the test image; (e) classification without illuminant compensation; (f) estimated illuminant distribution (black indicates outliers).

considered outliers. The assignment of illuminant types is shown in Figure 3 (d), while Figure 3 (e) shows the illuminant–compensated image; note how illuminant compensation "casts light" over the shaded areas. The classifier without illuminant compensation (Figure 3 (f)) finds several outlier pixels in the shadow area. If forced to make a choice without the outlier option (Figure 3 (g)), it misclassifies the pixels corresponding to stones in the pathway.

References

1. S. Buluswar and S.D. Draper, "Estimating Apparent Color in Outdoor Images", *Computer Vision and Image Understanding*, 85(2).
2. M. Tsukada and Y. Ohta, "An Approach to Color Constancy Using Multiple Images", *IEEE ICCV*, 385–389, 1990.
3. Y. Tsin, R.T. Collins, V. Ramesh and T. Kanade, "Bayesian Color Constancy for Outdoor Object Recognition", *IEEE CVPR*, December 2001.
4. D. Judd, D. MacAdam , G. Wyszecki, "Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature", *JOSA*, 54(8):1031–1040, 1964.
5. P. Bellutta, R. Manduchi, L. Matthies, K. Owens, A. Rankin, "Terrain Perception for DEMO III", *IEEE Intelligent Vehicles Symposium 2000*, Dearborn, MI, October 2000, 326-332.



(a)

(b) (c) (d)

(e) (f) (g)

Fig. 3. Color classification experiments: (a): training image; (b) test image; (c) classification with illuminant compensation (blue indicates outliers); (d) estimated illuminant distribution (black indicates outliers); (e) illuminant–compensated version of the test image; (f) classification without illuminant compensation; (g) classification without illuminant compensation (without outlier class).

6. D.H. Brainard and W.T Freeman, "Bayesian Color Constancy", *JOSA–A*, 14:1393–1411, July 1997.
7. M. D'Zmura and G. Iverson, "Color Constancy I - Basic Theory of Two–Stage Linear Recovery of Spectral Descriptions for Lights and Surfaces", *JOSA*, 10(10:2148–2165, October 1993.
8. G. Finlayson, M. Drew and B. Funt, "Diagonal Transforms Suffice for Color Constancy, *IEEE ICCV*, 162–171, 1993.
9. D. Forsyth, "A Novel Approach for Color Constancy", *International Journal of Computer Vision*, 5:5–36, 1990.
10. G. Healey, "Segmenting Images Using Normalized Colors", *IEEE Trans. Syst., Man and Cybern.*, 22(1):64–73, January 1992.
11. L.T. Maloney and B. Wandell, "A Computational Model of Color Constancy", *JOSA*, 1(1):29–33, January 1986.

12. C. Rosenberg, M. Hebert amd S. Thrun, "Color Constancy Using KL–Divergence", *IEEE ICCV*, 2001.
13. G.J McLachlan and T. Krisnan, *The EM Algorithm and Extensions*, Wiley, 1997.
14. W. H. Press, B. P. Flannery, S. A. Teukolsky andW. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambidge Unvierisity Press, 1986.
15. Y. Weiss and E.H. Adelson, "A Unified Mixture Framework for Motion Segmentation: Incorporating Spatial Coherence and Estimating the Number of Models", *IEEE CVPR*, 321–326, 1996.
16. J. Zhang, J.W. Modestino and D.A. Langan, "Maximum–Likelihood Parameter Estimation for Unsupervised Stochastic Model–Based Image Segmentation", *IEEE Trans. Image Proc.*, 3(4):404–420, July 1994.
17. J. Besag,"On the Statistical Analysis of Dirty Pictures", *J. Royal Stat. Soc. B.*, 48:259–302, 1986.
18. B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
19. H. Riad and R. Mohr, "Gaussian Mixture Densities for Indexing of Localized Objects in Video Sequences", INRIA Technical Report RR-3905, 2000.
20. E.T. Jaynes, *Probability Theory as Extended Logic*, unpublished, available at http://bayes.wustl.edu.

Appendix A

In this Appendix, we show how the total likelihood $L_F(C)$ can be maximized over the parameters $\{d_{l,m}\}$ using Expectation Maximization (EM). Using the diagonal illumination model, we can rewrite $L_F(C)$ as

$$L_D(C) = \sum_x \log \sum_{l=1}^{L} \sum_{k=1}^{K} P_K(k) P_L(l) p(D_l c(x)|k) \det |D_l| \tag{9}$$

As customary with the EM procedure, we first introduce the "hidden" variables $z_{l,k}(x)$ which represent the (unknown) label assignments: $z_{k,l}(x)$=1 if the pixel x is assigned to the illuminant type l and surface class k; $z_{k,l}(x)$=0 otherwise. We will denote the set of $z_{k,l}(x)$ over the image by Z.

The EM algorithm starts from arbitrary values for the diagonal matrices $\{D_l^0\}$, and iterates over the following two steps:

- Compute
$$Q\left(\{D_l\}, \{D_l^0\}\right) = E_{\{D_l^0\}}\left[\log p_{\{D_l\}}(C, Z)|C\right] \tag{10}$$
$$= E_{\{D_l^0\}}\left[\log p_{\{D_l\}}(C|Z)|C\right] + E_{\{D_l^0\}}\left[\log p_{\{D_l\}}(Z)|C\right]$$
where $E_{\{D_l^0\}}[\cdot|C]$ represents expectation over $p_{\{D_l^0\}}(Z|C)$.
- Replace $\{D_l^0\}$ with arg max $Q\left(\{D_l\}, \{D_l^0\}\right)$.

Given the assumed independence of label assignments, and the particular form chosen for variables $z_{k,l}(x)$, we can write

$$\log p_{\{D_l\}}(C|Z) = \sum_x \sum_{k=1}^{K} \sum_{l=1}^{L} z_{k,l} \log p(D_l c(x)|k) \tag{11}$$

where, given that the conditional likelihood are assumed to be Gaussian:

$$\log p(D_l c(x)|k) = -1.5\ \log(2\pi) - \log|\det \Sigma_k| \tag{12}$$
$$-0.5\ (D_l c(x) - \mu_k)\Sigma_k^{-1}(D_l c(x) - \mu_k)$$

Also,

$$\log p_{\{D_l\}}(Z) = \sum_x \sum_{k=1}^{K} \sum_{l=1}^{L} z_{k,l} \left(\log P_K(k) + \log P_L(l)\right) \tag{13}$$

It is easy to see that

$$E_{\{D_l^0\}}\left[\log p_{\{D_l\}}(C|Z)|C\right] = \sum_x \sum_{k=1}^{K} \sum_{l=1}^{L} \log p(D_l c(x)|k) P_{\{D_l^0\}}(k,l|c(x)) \tag{14}$$

and

$$E_{\{D_l^0\}}\left[\log p_{\{D_l\}}(Z)|C\right] = \sum_x \sum_{k=1}^{K} \sum_{l=1}^{L} \left(\log P_K(k) + \log P_L(l)\right) P_{\{D_l^0\}}(k,l|c(x)) \tag{15}$$

where $P_{\{D_l^0\}}(k,l|c(x)) = E_{\{D_l^0\}}[z_{k,l}(x)]$ is the posterior probability of surface class k and illumination type l given the observation $c(x)$ and under color transformation matrices $\{D_l^0\}$. Using Bayes' rule, we can compute $P_{\{D_l^0\}}(k,l|c(x))$ as

$$P_{\{D_l^0\}}(k,l|c(x)) = \tag{16}$$
$$\frac{P_K(k)P_L(l)p(D_l^0 c(x)|k)\det|D_l^0|}{\sum_{\bar{k}=1}^{K}\sum_{\bar{l}=1}^{L} P_K(\bar{k})P_L(\bar{l})p(D_{\bar{l}}^0 c(x)|\bar{k})\det|D_{\bar{l}}^0|}$$

Remembering that $|D_l| = |d_{l,1}d_{l,2}d_{l,3}|$ one sees that

$$\max_{\{d_{l,m}\}} Q\left(\{D_l\},\{D_l^0\}\right) \tag{17}$$
$$= \max_{\{d_{l,m}\}} \sum_{l=1}^{L}\left[u_l \sum_{m=1}^{3} \log|d_{l,m}| - 0.5\ d_l' G_l d_l + H_l' d_l\right]$$

where $d_l = (d_{l,1}, d_{l,2}, d_{l,3})'$, and

$$u_l = \sum_x \sum_{k=1}^{K} P_{\{D_l^0\}}(k,l|c(x)) \tag{18}$$

$$G_{l,m,s} = \sum_x \sum_{k=1}^{K} S_{k,m,s} c_m(x) c_s(x) P_{\{D_l^0\}}(k,l|c(x)) \tag{19}$$

$$H_{l,m} = \sum_x \sum_{k=1}^{K} \sum_{s=1}^{3} S_{k,m,s} c_s(x) \mu_{k,s} P_{\{D_l^0\}}(k,l|c(x)) \tag{20}$$

where $S_k = \Sigma_k^{-1}$. Note that the terms in the summation over l in (17) can be maximized independently, meaning that the L sets of diagonal transformations can be computed independently at each iterations.