Object-Based Illumination Classification

H. Z. Hel-Or

Dept. of Computer Science
Haifa University
Haifa 31905, Israel

B. A. Wandell

Dept. Of Psychology
Stanford University
Stanford, CA 94305, USA

Abstract

Estimation of Scene Illumination from a single image or an image sequence has been widely studied in computer vision. The approach presented in this paper, introduces two new issues: 1) Illumination classification is performed rather than illumination estimation. 2) An object based approach is used for Illumination Evaluation. Thus pixels associated with an object are considered in the illumination estimation process using the object's spectral characteristics. Simulation and real image experiments, show that the Object Based approach indeed improves performance over standard Illumination classification.

Keywords: Illumination estimation, Color correction, Face, Color, Classification

1 Introduction

Scene estimation is a basic problem in image understanding: given an image obtained from an acquisition system, estimate the spectral distribution of the scene illuminant. A classic model for image acquisition states that light from a source impinges on a surface and is partially reflected from it. The reflected light is measured by the camera sensors and the acquired image is produced from the sensor outputs. The light source is given as a wavelength

dependent intensity function - the *spectral distribution function* (SPD) $E(\lambda)$. The surface reflectance is the proportion of incident light which is reflected as a function of spectral wavelength, and is given by the *surface reflectance function* $S(\lambda)$ (with values between 0 and 1). The surface reflectance function is characteristic of the surface material and is associated with the color appearance of the surface. The light reflected from the surface is denoted the *color signal* and is given by (assuming a Lambertian model):

$$C(\lambda) = E(\lambda)S(\lambda) \tag{1}$$

The camera sensors, which measure the color signal, supply a single output computed as a weighted integral over the spectral wavelengths of the color signal. The weights correspond to the sensor sensitivity at each spectral wavelength. There are typically three types of sensors commonly referred to as R, G and B corresponding to their spectral sensitivities. The acquired image is composed of pixels, each of which contains the output values of the camera sensors at the pixel position. The pixel value p_i corresponding to the surface reflectance $S(\lambda)$ viewed under illuminant $E(\lambda)$ and acquired by the i-th camera sensor $R_i(\lambda)$ is given by:

$$p_i = \int_\lambda C(\lambda)R_i(\lambda)d\lambda = \int_\lambda E(\lambda)S(\lambda)R_i(\lambda)d\lambda \tag{2}$$

where the integration is performed over the visible spectrum (electromagnetic radiation at wavelengths approximately 400nm to 700nm).

Equation 2 provides the basis for Illumination Estimation and Illumination Classification algorithms. By sampling (using N_λ samples) the SPDs associated with the illuminant and the surface reflectances, Equation 2 can be rewritten in vector notation as:

$$\mathbf{p_j} = \mathbf{RES_j} \tag{3}$$

where $\mathbf{p_j}$ is the 3-vector of sensor outputs at pixel location j, S_j is the surface reflectance corresponding to pixel location j represented as an $N_\lambda \times 1$ vector, $\mathbf{E}$ is the illuminant of the scene represented as a $N_\lambda \times N_\lambda$ matrix with the sampled $E(\lambda)$ on the diagonal, and $\mathbf{R}$ is a matrix with rows corresponding to the sampled camera sensor sensitivities.

The problem of illumination estimation is to determine the SPD $E(\lambda)$ of the scene illuminant given the image pixel values. The difficulty in estimating the illuminant originates from the

confounding nature of the color signal: many illuminant-surface pairs can explain a given color signal. Additionally, the acquisition system produces sensor outputs and not SPDs, thus metameric pairs (color signals producing the same sensor output) add to the confusion. The importance of estimating scene illumination is demonstrated in Figure 1. A surface is viewed under two different illuminations. Following Equation 1, the color signal is calculated from the surface reflectance function (Figure 1a) and the illuminant SPD (Figures 1b and 1c). The two color signals (Figures 1d and 1e) are then measured by the camera sensors and the sensor responses (Figures 1f and 1g) are computed using Equation 2. As can be seen, the color signal, and accordingly the sensor outputs, produce two different responses for the same surface. However the Human Visual System (HVS) is inconsistent with this behavior. A characteristic of the HVS is to correct for the illumination and produce a stable color appearance regardless of illumination. This characteristic is known as *Color Constancy*. Thus, an apple appears red whether it is viewed under red-biased indoor tungsten illumination or blue-biased sunlight illumination. Given that observers demand realistic reproductions of a scene (as determined by the HVS), photo developers and printers must correct the effects of the illumination before producing the final image (else an outdoor image of an apple will have a blue tint). This correction is known as *Color Correction*. With the introduction of digital cameras, in which images are directly viewed on the computer monitor following acquisition, color correction must be performed automatically in the camera or driver. This promotes the need for efficient algorithms for scene illumination estimation.

Various approaches to Illumination Estimation have been previously suggested. Typically the image acquisition model is considered for each pixel, supplying a single equation (Equation 2). The set of equations is then solved for the illuminant and the reflectance functions. Usually, this system of equations is under-determined so that additional constraints and assumptions must be made. Previous approaches are distinguished by the assumptions and constraints imposed on the image, scene or acquisition model [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12].

2 Our Approach

In estimating scene illumination, the fact that color values of an acquired image originate from a natural scene, must be taken into account. We suggest that information about the scene obtained from the acquired image will greatly increase performance of the illumination estimation process. In other words, illumination estimation is dependent on the objects appearing in the given image. This approach is strongly supported by the fact that color constancy has been shown to improve when well known objects appear in the scene (compared to scenes of color patches etc) [13].
Our approach ([14]) introduces two new issues:

1. Illumination classification is performed rather than illumination estimation.

2. Illumination evaluation is object dependent.

Illumination Classification

Rather than estimating the exact spectral distribution of the scene illumination, we perform *Illumination Classification*. The problem then reduces to: Given a finite set of illuminant classes, determine which of the classes is most probable as the illuminant of a given image. Classification is sufficient for color correction since efficient algorithms use look-up tables (rather than closed form solutions or computations). The look-up table to be used is determined by the estimated scene illuminant. A finite number of look-up tables are used in the process, corresponding to a finite number of illuminant classes. Illumination Classification rather than Illumination Estimation, simplifies the problem at hand: the exact illuminant need not be recovered. The solution space is finite and the solution becomes more robust.

Content Dependent Illumination Estimation

Previous approaches to illumination estimation consider the sensor outputs as a set of values, irrespective of the pixel source. The estimation process considers the sensor outputs at each pixel as contributing equally to the final solution. In our approach, subsets of pixel values are considered differently. Specifically, subsets of pixels are extracted or segmented from

the image and the associated sensor outputs are exploited more efficiently in the estimation process. As will be discussed in Section 4, the estimation process we assume is *Object Based*. Thus, pixels associated with an object are considered in the illumination estimation process using the object's spectral characteristics.

3 Illumination Classification

Given an illuminant, we consider the set of all surface reflectances viewed under the illuminant by a known set of camera sensors. A large set of possible sensor outputs is obtained forming a cluster of points in the 3D sensor output space. Given a finite number of illuminants, several such clusters are obtained, one for each illuminant. Figure 2a shows two Illuminant Clusters obtained by plotting the Munsell surfaces as imaged by the three Kodak DCS-20 camera sensors under two illuminants - tungsten bulb (CIE illuminant A) and typical daylight (CIE illuminant C).

An image of a scene, is assumed to be a subset of the set of surfaces imaged by the camera sensors under one of the illuminants. As such, the image pixels form a cluster of points in 3D which is a subset of one of the illuminant clusters defined above. Thus, given the cluster of 3D points of the image pixel values, the scene illumination can be determined by classifying the image cluster to one of the illuminant clusters.

Intuitively, the image cluster should be classified to the "closest" illuminant cluster. Some metric must be used to measure the "distance" between the image cluster and each illuminant cluster. Given several illuminant clusters, it is obvious that the greater the "distance" between the clusters, the better the illumination classification will perform.

3.1 Cluster Distance Metric

The distance metric used in this work is the *Mahanalobis Distance* [15] which is defined between two Gaussian distributed clusters as:

$$d = (M_1 - M_2)^t(\Lambda_1 + \Lambda_2)^{-1}(M_1 - M_2)$$

Where M_i and Λ_i are the mean vector and the covariance matrix of cluster i, respectively. As a degenerate case this metric can also measure the distance between a single 3D point P and a cluster:

$$d = (P - M_1)^t \Lambda_1^{-1} (P - M_1)$$

Note that the Mahanalobis distance decreases with the decrease in distance between the means and with the increase in cluster covariance.

Using this metric, several statements can be deduced from the simple simulation of illumination of Munsell surfaces (see Figure 2b). Given the two reflectance clusters of Figure 2a, the Mahanalobis distance between the cluster means is 1.79, whereas the distance between the means of the 30% brightest points is 5.57 and the 30% darkest points is 1.41. Intuitively, this is reasonable: The dark surfaces reflect very little light so that discrimination between the clusters is difficult. The bright surfaces, on the other hand, reflect a large percentage of the illuminant. Since the illuminants are easily distinguishable, so are the brightest points and, accordingly, classification is efficient [16]. This can also explain the motivation behind several approaches to illumination estimation where only bright pixels are considered and dark pixels are rejected.

The brightest colors in an image and the mean color of the image do not always produce satisfactory illumination estimation results since the brightest pixels in the scene may not be bright enough to distinguish between the illuminant clusters. Similarly when considering the mean of all colors in the image, the large variability (large spread of the image cluster), reduces the Mahanalobis distance to all illuminant clusters, producing unreliable classification.

As mentioned above, to improve performance of illumination classification, the distance between illuminant clusters should be maximized. In the following section we describe how this is performed using the Object Based illumination estimation method. A different approach to increasing inter-cluster distance, is by varying the camera sensors, which essentially varies the projection of the high dimensional color signal space to the lower dimensional camera sensor space. Using the cluster notation, optimal filters for illumination classification can be calculated. This approach is outside the scope of this paper and will be described elsewhere.

4 Object Dependent Illumination Classification

Previous approaches to illumination estimation consider the sensor outputs provided by the image as a set of values. Each pixel location j in the image provides an equation of the form defined by Equation 3. The origin (in terms of spatial location in the image) of each equation is lost, and the estimation process considers each equation, and accordingly each pixel in the image, as contributing equivalently to the final solution (Figure 3a).

In our approach, subsets of pixel values are considered differently. Subsets of sensor outputs are exploited more efficiently and provide additional information for the Illumination Classification process (Figure 3b). The subset approach suggested in this section provides better classification as follows:

The set of pixel values, $\mathbf{p_j}$, extracted from an image is a subset of one of the illumination clusters (e.g. Figure 2a). That is, each $\mathbf{p_j}$ can be any point in the illumination clusters, since the corresponding surface reflectance $\mathbf{S_j}$, may be any one of the surface reflectances used to create the illumination cluster. However, given apriori spectral information about the surface reflectance $\mathbf{S_j}$, the set of points of the illuminant cluster from which $\mathbf{p_j}$ may originate is constrained. Figure 4 shows two Illumination Clusters obtained as described in Section 3. Previous approaches assume $\mathbf{p_j}$ may originate from any one of the points in the clusters. Apriori knowledge on $\mathbf{S_j}$ constrains $\mathbf{p_j}$ to originate from a subset of cluster points, marked with an ellipse for each Illumination Cluster. If the distance between these subsets of cluster points is greater than the distance between the Illumination Clusters themselves, then classifying using the apriori knowledge on $\mathbf{S_j}$ provides better classification.

4.1 Object-Based Illumination Classification

In order to provide apriori spectral information on the surface reflectance $\mathbf{S_j}$ associated with the pixel values $\mathbf{p_j}$, we use an *Object Based* approach. Thus using spatial information in the image, pixels associated with an object can be segmented, extracted and used in the illumination estimation process while exploiting the object's spectral characteristics.

Most classes of material are constrained in terms of their spectral reflectance characteristics. Associating certain material with objects found in the scene provides the apriori spectral in-

formation that we exploit. For example, sky, foliage, grass, skin etc have constrained surface reflectance characteristics. Computer Vision techniques can be used to find object pixels in an image on which the Object Based Illumination Classification technique can be applied. In this paper we examine the use of Human Skin spectral reflectance to improve Illumination Classification. Specifically, pixels associated with faces are used in the classification process. The choice of faces as the objects for our technique was made for the following reasons. First, a large proportion of photographs and images contain faces. Additionally, the relatively small variability in skin tones (see Section 5) produces a tight apriori constraint on the pixels which is advantageous for the classification . Finally, being such a dominant object in our visual world and in images specifically, many Computer Vision techniques have been developed to detect faces in images (e.g. [17, 18, 19, 20, 21]). These face detection tools have reached a relatively high rate of success, promising successful automatic detection of face pixels.

5 Results

The spectral reflectance characteristics of Human Skin has been studied and measured. We consider the skin spectral reflectance data given in [22] and shown in Figure 5a. Figure 5b shows the distribution of all surface reflectances (Munsell) vs skin surface reflectances as viewed under CIE illuminant A (tungsten) using the Kodak DCS-200 camera sensors. The points in the cluster associated with the skin surfaces form a small subset. Given two Illuminations (CIE-A and CIE-C) The Mahanalobis distance between the Illumination Clusters associated with all surface reflectances is shown in Figure 6a. The Mahanalobis distance between the Illumination Clusters associated with the skin surfaces is shown in Figure 6b. As expected, the inter-cluster distance is greater between the Illumination Clusters associated with the skin surfaces.

Illumination Classification - Simulation

Four illuminants were chosen (Figure 7a). For each illuminant, the Illumination Clusters

associated with all Munsell surface reflectances (Figure 7b) and skin surface reflectances (Figure 7c) was created.

The simulation was performed as follows: Given an image containing a human face which was acquired under white illumination (Figure 8a), we chose one of the four illuminants, added zero mean Gaussian noise to the illuminant and created a new image containing the same scene as the original but viewed under the new illuminant (an example is shown in Figure 8b). The pixel values of the new image were then used to determine the scene Illuminant by classifying according to the four original illuminant clusters using all image pixels and using only the skin pixels which were extracted manually. The process was repeated 40 times (10 times for each illuminant). Simulation results are given in Table 1.

As expected, Illumination Classification according to skin pixels out performs classification according to all image pixels with no apriori spectral information.

Illumination Classification - Real Images

A set of 12 calibrated images containing a human face was obtained under 8 different illuminants. The illuminants of the scene were measured for each acquired image. Figure 9 shows several of the images and their scene illuminants. The 8 illuminant clusters associated with all surface reflectances and associated with the skin surface reflectances are shown in Figure 10a and 10b respectively. Illumination Classification was applied to the 12 images using all image pixels and using the skin pixels (which were manually extracted). Results of the classification are shown in Table 2.

Results show that Illumination Classification according to skin pixels out performs classification according to all image pixels with no apriori spectral information.

6 Conclusion

In this paper a new approach to Illumination Estimation has been suggested. We have considered how one might integrate illuminant estimation and object estimation. In general, pixels associated with an object will provide a more secure basis for illuminant estimation. Given that it should prove possible to regularly identify faces in images, and that faces fre-

quently are present in images, we evaluated how much improvement might be expected from performing illuminant estimation based on human skin. Simulation and real image experimentation showed an advantage for such an object-based illuminant estimation approach.
The results presented, deal with skin surfaces, however other classes of pixels can be used (foliage, sky, etc). Apriori information on the scene can similarly be exploited without considering objects in the scene (for example specific scene types such as agricultural fields, city scenes involving roads and pavements etc). Finally, if no pixels are found with which apriori information can be associated, the technique reverts back to the simple classification using voting. Thus, the Object Based Illumination Classification technique can be viewed as a general classification technique with weighted votes. Votes arising from pixels associated with apriori information are weighted more heavily than other votes.
A parameter that was not dealt with in this paper, nor in previous studies, is the number of Illumination Classes. For color correction, evaluation of the sensitivity of the HVS to changes in scene illumination is necessary to determine the number of Illumination Classes and, accordingly, the number of LUTs to be used in the correction process. The more sensitive the HVS, more Illumination Classes must be used, and the classification becomes more difficult, increasing the need for approaches such as the Object Based approach suggested in this paper. This evaluation is outside the scope of this paper.

Acknowledgment The authors thank Joyce Farell and HP Labs, for providing the calibrated images.

References

[1] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Optical Society of America A*, 61:1–11, 1971.

[2] G. Buchsbaum. A spatial processor model for object color perception. *J. Franklin Inst.*, 310:1–26, 1980.

[3] M. D'Zmura and P. Lennie. Mechanisms of color constancy. *J. Optical Society of America A*, 3(10):1662–1672, 1986.

[4] L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *J. Optical Society of America A*, 3(1):29–33, 1986.

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

[6] B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image processing*, 3:277–299, 1974.

[7] H. C. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *J. Optical Society of America A*, 3(10):1694–1699, 1986.

[8] S. A. Shafer. A physical approch to color image understanding. *Int. J. of Computer Vision*, 4:210–218, 1990.

[9] W. T. Freeman and D. H. Brainard. Bayesian method for recovering surface and illuminant properties from photosensor responses. In *Proc. of the SPIE*, volume 2179, pages 364–376, 1994.

[10] D. A. Forsyth. A novel algorithm for color constancy. *Int. J. of Computer Vision*, 5:5–36, 1990.

[11] G. D. Finlayson, M. S. Drew, and B. V. Funt. Color constancy: Generalized diagonal transforms suffice. *J. Optical Society of America A*, 11(11):3011–3019, 1994.

[12] G. D. Finlayson, P. M. Hubel, and S. Hordley. Color by correlation. In *5th IS&T/SID Color Imaging Conference*, pages 6–11, Scottsdale, Arizona, 1997.

[13] D. I. Bramwell and A. C. Hurlbert. The role of object recognition in colour constancy. *Perception*, 22:S62, 1993.

[14] H.Z. Hel-Or and B.A. Wandell. Illumination classification based on image content. In *Annual Symposium of the Optical Society of America OSA-98*, Baltimore MA, Oct. 1998.

[15] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis.* Wiley Press, New York, 1973.

[16] S. Tominaga, S. Ebisui, and B. Wandell. Scene illumination classification: Brighter is better. *J. Optical Society of America A*, 18(1):55–64, 2001.

[17] K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20:39–51, 1998.

[18] H.A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20:23–38, 1998.

[19] A.N. Rajagopalan, K.S. Kumar, J. Karlekar, R. Manivasakan, and M.M. Patil. Finding faces in photographs. In *International Conference of Computer Vision (ICCV) 97, Delhi, India*, 1997.

[20] M. Elad, Y. Hel-Or, and R. Keshet. Pattern detection using a maximal rejection classifier. In *Fourth International Workshop on Visual Form*, Capri, Italy, May 2001.

[21] D. Keren, M. Osadchy, and C. Gotsman. Anti-faces: A novel, fast method for image detection. To Appear in *IEEE Trans. Pattern Analysis and Machine Intelligence.*

[22] G. Wyszecki and W. S. Stiles. *Color science: concepts and methods, quantitative data and formulae.* Wiley Press, New York, 1982.

Figure Captions

Figure 1: A single surface viewed under two different illuminations.

Figure 2: The set of Munsell surfaces imaged by three sensors under the two illuminants: tungsten (CIE-A) and typical daylight (CIE-C). a) For each illuminant, the set of surfaces forms a cluster of points in the 3D sensor output space (marked + and ∘ respectively). A 2D projection of the sensor space is shown (the R and B camera sensors). b) The Mahanalobis distance is computed between subsets of the cluster points shown in a): Between the two clusters d=1.79 (middle). Between the 30% brightest points of the two clusters d=5.57 (top). Between the 30% darkest points of the two clusters d=1.41 (bottom). The means of the subsets are represented as + and ∘. The spread of each subset is shown as an ellipse with width and length proportional to the covariance matrix. Insets show the illuminants and the camera sensors.

Figure 3: Two approaches to Illumination Classification.
a) Classic approach b) Image content dependent approach.

Figure 4: Two Illumination Clusters obtained for illuminants CIE-A and CIE-C. Classic approaches assume the sensor output originate from any one of the points in the clusters. Apriori knowledge on the surface constrains the associated sensor output to originate from a subset of cluster points, marked with an ellipse for each Illumination Cluster.

Figure 5: a) Spectral reflectance functions of Human Skin. b) Distribution of Munsell surface reflectances (∘) vs skin surface reflectances (+).

Figure 6: The Mahanalobis distance between: a) Illumination Clusters associated with all surface reflectances. b) Illumination Clusters associated with the skin surfaces.

Figure 7: Illumination Classification - Simulation.
a) Four illuminants. b) The Illumination Clusters associated with all Munsell surface reflectances. c) The Illumination Clusters associated with the skin surface reflectances.

Figure 8: a) A face image acquired under white illumination. b) The image in a) transformed so that the scene is illuminated under CIE-A.

Figure 9: Images under different scene illuminants.

Figure 10: The 8 illuminant clusters associated with: a) all Munsell surface reflectances b) the skin surface reflectances.

Mode of Classification	Number of Hits	Percentage of Hits
All Pixels	26	65%
Skin Pixels	37	92.5%

Table 1: Illumination Classification Results

True Illuminant Class	Classification using All Reflectances	Classification using Skin Reflectances
1	6	1
2	2	2
3	3	3
3	2	3
4	2	4
5	6	5
6	1	1
6	6	6
7	7	7
8	8	8
8	8	8
8	8	8
Errors:	5	1

Table 2: Illumination Classification Results

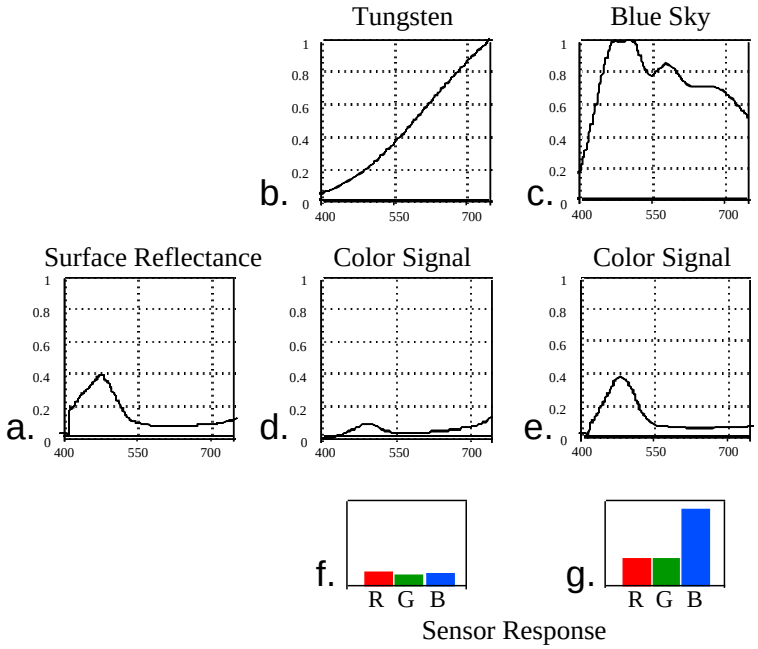

Figure 1: A single surface viewed under two different illuminations.

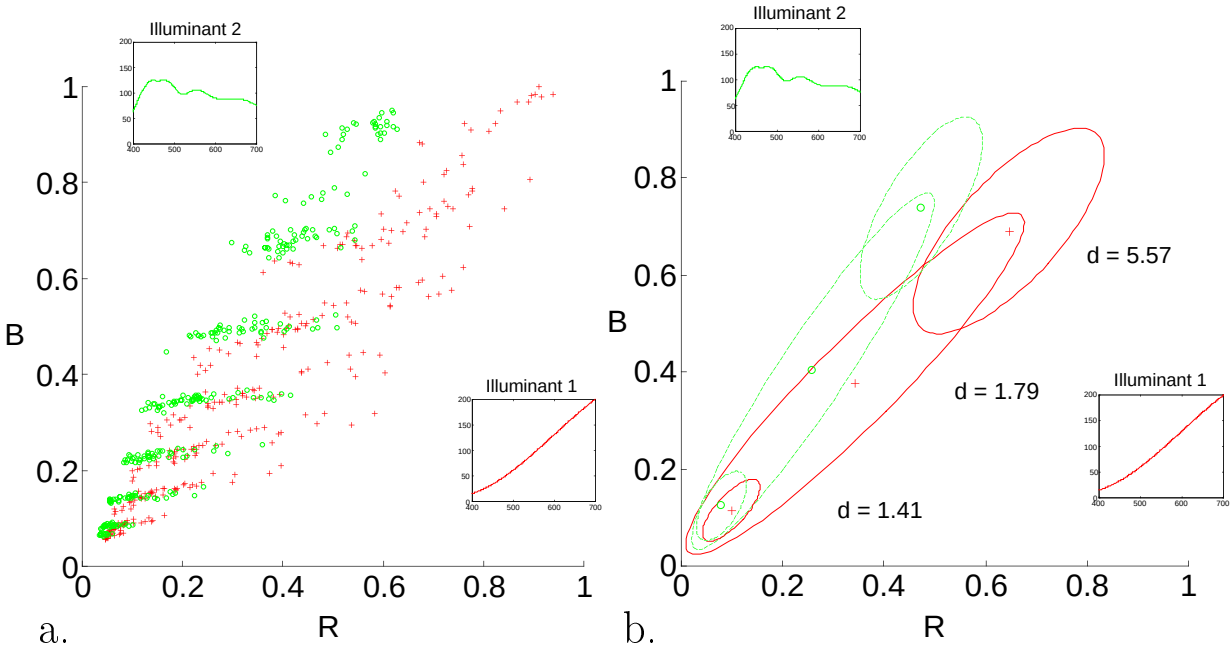

Figure 2: The set of Munsell surfaces imaged by three sensors under the two illuminants: tungsten (CIE-A) and typical daylight (CIE-C). a) For each illuminant, the set of surfaces forms a cluster of points in the 3D sensor output space (marked + and ∘ respectively). A 2D projection of the sensor space is shown (the R and B camera sensors). b) The Mahanalobis distance is computed between subsets of the cluster points shown in a): Between the two clusters d=1.79 (middle). Between the 30% brightest points of the two clusters d=5.57 (top). Between the 30% darkest points of the two clusters d=1.41 (bottom). The means of the subsets are represented as + and ∘. The spread of each subset is shown as an ellipse with width and length proportional to the covariance matrix. Insets show the illuminants and the camera sensors.

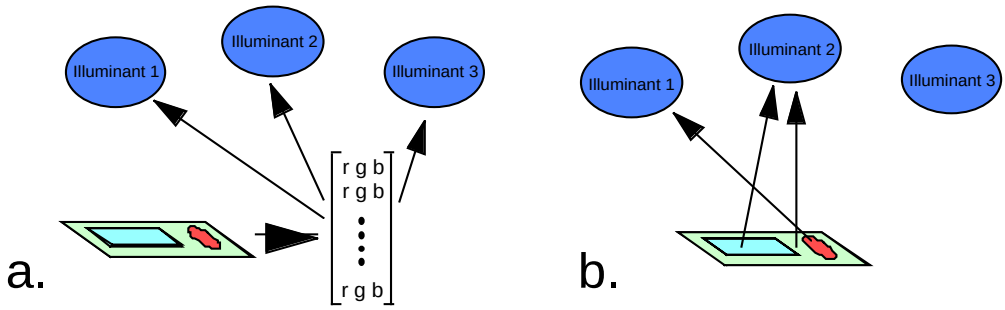

Figure 3: Two approaches to Illumination Classification. a) Classic approach b) Image content dependent approach.

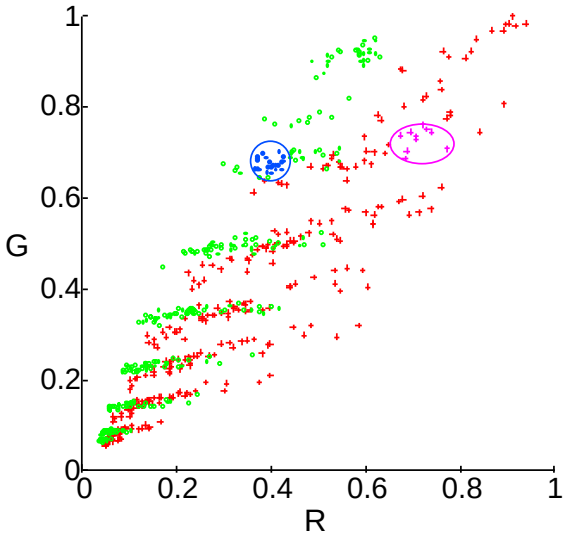

Figure 4: Two Illumination Clusters obtained for illuminants CIE-A and CIE-C. Classic approaches assume the sensor output originate from any one of the points in the clusters. Apriori knowledge on the surface constrains the associated sensor output to originate from a subset of cluster points, marked with an ellipse for each Illumination Cluster.

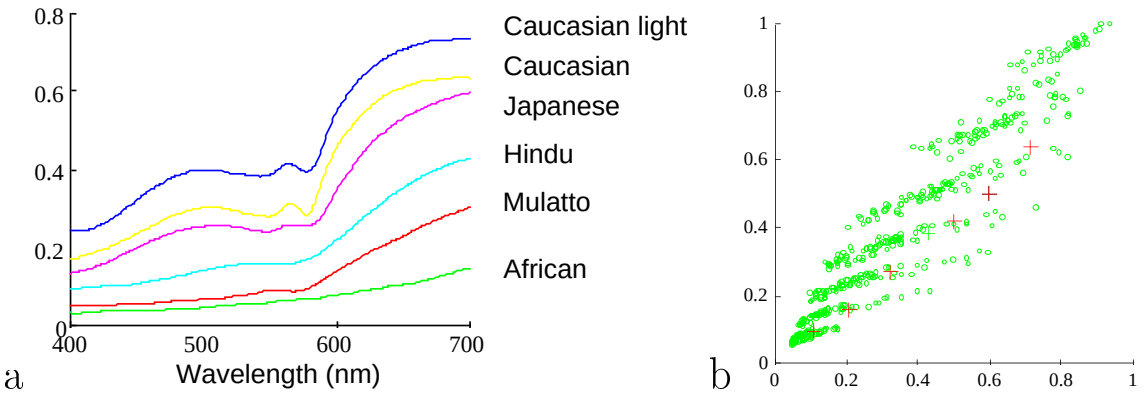

Figure 5: a) Spectral reflectance functions of Human Skin. b) Distribution of Munsell surface reflectances (∘) vs skin surface reflectances (+).

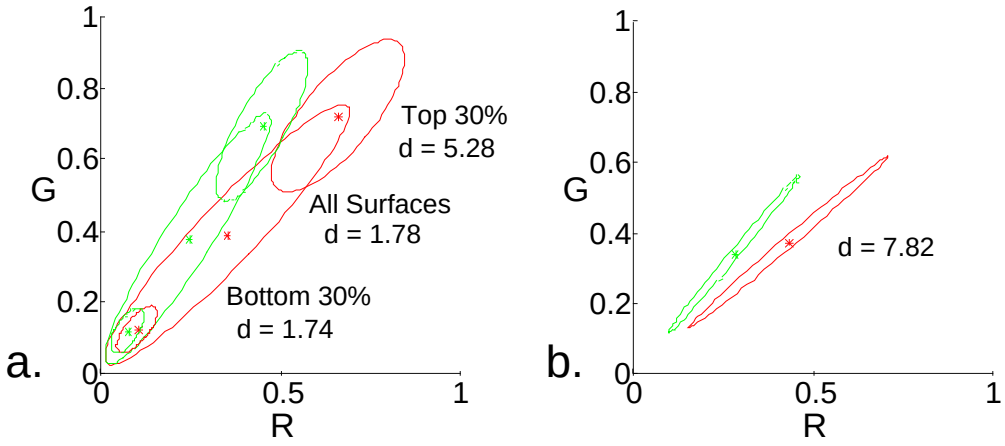

Figure 6: The Mahanalobis distance between: a) Illumination Clusters associated with all surface reflectances. b) Illumination Clusters associated with the skin surfaces.

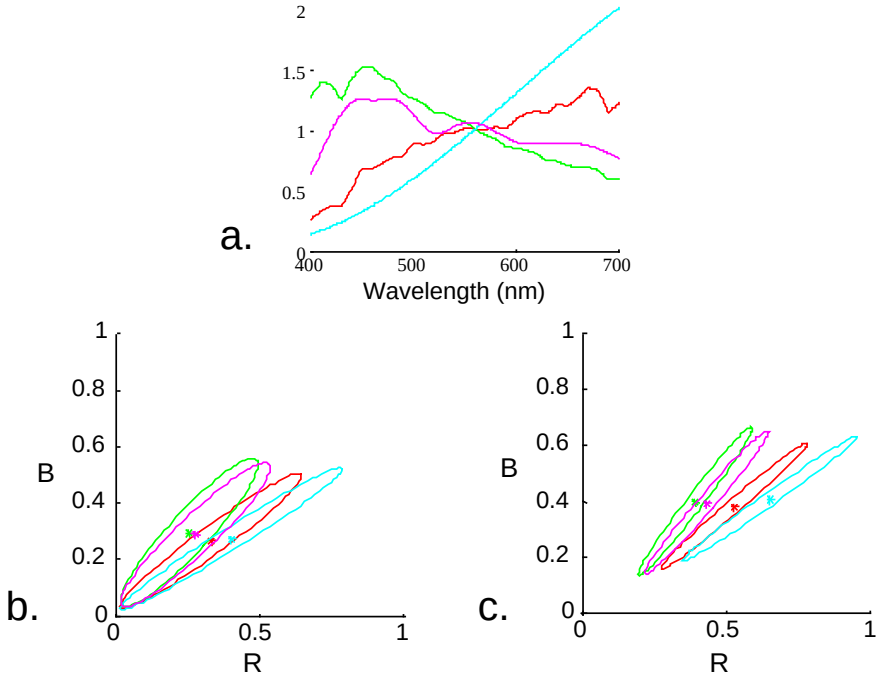

Figure 7: Illumination Classification - Simulation.

a) Four illuminants. b) The Illumination Clusters associated with all Munsell surface reflectances. c) The Illumination Clusters associated with the skin surface reflectances.

a.  b. 

Figure 8: a) A face image acquired under white illumination. b) The image in a) transformed so that the scene is illuminated under CIE-A.

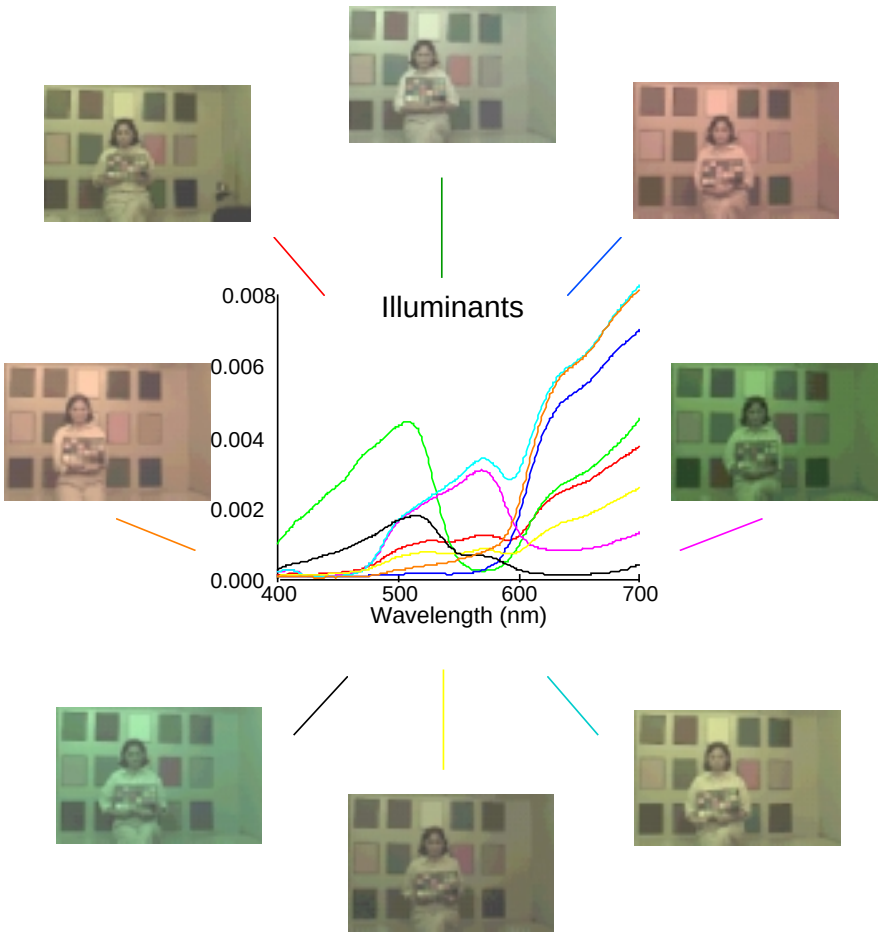

Figure 9: Images under different scene illuminants

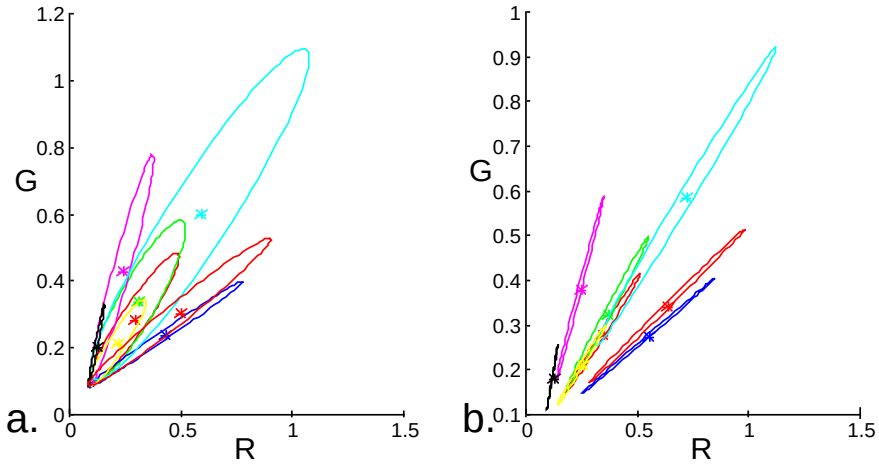

Figure 10: The 8 illuminant clusters associated with: a) all Munsell surface reflectances b) the skin surface reflectances.