Detecting and Removing Specularities and Shadows in Images

By

Jisnu Bhattacharyya

Electrical and Computer Engineering
McGill University, Montreal

September 2004

A thesis submitted to McGill University
in partial fulfillment of the requirements of the Degree of
Master of Engineering

Your file Votre référence
ISBN: 0-494-06544-3
Our file Notre référence
ISBN: 0-494-06544-3

Abstract

Specularities and shadows often confound algorithms designed to solve computer vision tasks such as image segmentation, object detection and tracking. In this thesis, an illumination compensation technique that detects and removes both specularities and shadows is proposed. The method requires no camera calibration or other a priori information regarding the scene. This thesis also introduces two new illumination invariant representations based on the Retinex: the $\widehat{R}$ image and the $\widehat{R}_{EDGE}$ image.

Potential specularities are initially detected and a wavefront grown outwards from the center of the specularity. This continues until the specularity boundary or a material boundary is reached. The latter is detected by the newly discovered illumination invariant $\widehat{R}$ and $\widehat{R}_{EDGE}$ images that are introduced in this thesis. Upon reaching the specularity boundary, the wavefront contracts inwards, coloring in the specularity as it contracts, until the specularity no longer exists.

After treating specularities, shadows are dealt with. Support Vector Machines are trained to identify shadow boundaries based on their boundary properties. This boundary information is used to identify shadowed regions in the image and then assign them the color of non-shadow neighbors of the same material.

Illumination compensation as proposed in this thesis was found to increase the accuracy of image segmentation, skin detection and face recognition.

Résumé

Les ombres et les spéculaires confondent souvent des algorithmes conçus pour résoudre des tâches de vision par ordinateur telles que la segmentation d'image, la détection d'objet et la poursuite d'objet mobile. Dans cette thèse, on propose une technique de compensation d'illumination qui détecte et enlève des spéculaires et des ombres. La méthode n'exige aucun calibrage d'appareil-photo ou toute autre information a priori concernant la scène. Cette thèse présente également deux représentations basées sur le Retinex qui sont invariables aux changements d'illumination: l'image $\widehat{R}$ et l'image $\widehat{R}_{EDGE}$.

Des spéculaires potentiels sont au commencement détectés et un front des ondes s'augmente vers l'extérieur du centre du spéculaire. Ceci continue jusqu'à la frontière spéculaire ou une frontière matérielle est atteinte. Le dernier est détecté par les images $\widehat{R}$ et $\widehat{R}_{EDGE}$ nouvellement découvertes et présentés dans cette thèse. Lors d'atteindre la frontière spéculaire, le front des ondes se contracte vers l'intérieur, colorant le spéculaire tout en se contractant, jusqu'à ce que le spéculaire n'existe plus.

Après avoir traiter les spéculaires, les ombres sont traitées. Des Support Vector Machines sont formées pour identifier des frontières d'ombre basées sur leurs propriétés de frontière. Cette information de frontière est employée pour identifier des régions ombragées dans l'image et puis pour leur assigner la couleur des voisins non-ombragées du même matériel. La compensation d'illumination comme proposée dans cette thèse peut augmenter l'exactitude de la segmentation d'image, de la détection de peau et de l'identification de visage.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Levine, for without his inspiration, patience and vast knowledge this thesis would not have been possible. From the beginning, he was always available to help me when difficulties arose. Thanks to his guidance, understanding and expertise I was able to successfully complete this challenging undertaking.

My wife Victoria deserves an award for the vast amounts of encouragement she gave me during graduate school and during the writing of this thesis. She never once wavered in her support, and I could not have finished this thesis without her. I am grateful for the understanding of my parents and family over the last two years as I finished my Master's, especially during those times when the research came first. My father deserves thanks for making me believe that I could always do whatever I put my mind to as long as I worked hard enough. He would have been very proud of me. And, of course, I am grateful to my precious newborn son Indra Kumar and my affectionate cats for always reminding me to appreciate the priceless little moments every day that often get washed away in the tide of life.

Finally, I would like to thank my fellow graduate students – Bhavin Shastri, Maulin Gandhi, Gurman Gill and Ajit Rajwade – for not only helping me in technical matters, but for the friendship and good times we shared. May their camaraderie and unbridled optimism be noted in the annals of graduate history.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Need for Illumination Compensation

From a practical point of view, variations in illumination due to shadows, specularities and poor lighting can cause problems for computer vision algorithms such as segmentation, tracking, or object recognition. Due to discontinuities resulting from variable or poor illumination, a given material may be segmented into several regions, as illustrated in Figure 1.1. What is required is illumination compensation - correction for the impact of these illumination differences – so that subsequent computer vision algorithms can deliver increased accuracy based on uniform material properties alone.

Figure 1.1 Image segmentation. **Left:** Original, **Right:** Segmented image. Segmentation is discussed in more detail in Section 4.5.

From a theoretical point of view, the light reflected from a surface is the product of its reflectance and the illumination of the scene [1,2]. If the intensity and spectral distribution of the the *illumination* at each point in an image is known, then the reflectance can be recovered. However, the intensity of the illumination will vary according to the geometry of the scene, the angle of incidence of the illuminant and the viewing angle. A priori knowledge of all these factors is possible in a very controlled laboratory setting, but such information is not available for typical images. A method that compensates for illumination without imposing unnatural constraints is proposed in this thesis.

The proposed method, which only requires that scenes be illuminated by common indoor and outdoor illuminants [3,4,5], first uses the Luminance Retinex [6] - a variant of the Retinex [7-15] - to correct for poor illumination in an image. Specularities are then detected by using a technique motivated by [16]. Each specularity is removed by growing outward from the center of the specularity to its boundary, or until a material change is detected . Next, shadows are detected by segmenting the image into regions and using a Support Vector Machine to identify the boundaries of shadows. Once the boundaries of shadows have been detected, the corresponding shadows are given the color of their non-shadow neighbors (of the same material).

The final illumination compensated image enjoys good color uniformity for materials and is remarkably free of specularities and shadows. An example is shown in Figure 1.2.

(a) (b)

Figure 1.2 Segmentation after illumination compensation. **Left:** Left-most image from Figure 1.1 after illumination compensation, **Right:** Illumination compensated image after segmentation. Compare with Figure 1.1.

However, these images do not fare exceptionally well for the purposes of image enhancement as they seem somewhat artificial at times. Nevertheless, the benefits are excellent for the purposes of computer vision algorithms such as segmentation and object recognition.

1.2 Background and Literature Review

Pre-processing algorithms for illumination compensation include general image processing tools such as the traditional histogram equalization and gamma correction which correct for poor lighting effects by modifying the dynamic range of an image. The so-called Retinex [7-15] is an image enhancement and illumination compensation algorithm that also modifies the dynamic range of a scene while additionally offering a considerable amount of color constancy. The Retinex is discussed in detail in Chapter 2.

Traditional pre-processing algorithms do not specifically treat strong specularities and shadows. Classical approaches to detecting and removing specularities make use of the dichromatic reflection model [17], which is discussed in depth in Chapter 4. These methods [17-23] are quite successful in controlled settings with uniform backgrounds and foreground objects that have very saturated colors, such as plastic spheres. Analyzing and removing highlights in images with complex scenes has proven more successful when photometric stereo [19,24-27] methods have been used along with the dichromatic reflection model. In this thesis photometric stereo techniques do not concern us as the focus is on removing the highlights given a single image. Recently, however, Torres [16] has shown that potential specularities can be thresholded in a single image by using a binary mask in intensity-saturation space. We use this idea as the basis for our specularity algorithn which is discussed in detail in Chapter 4.

Several algorithms for shadow detection and removal exist in the literature and are discussed in depth in Chapter 5. These methods usually impose unnatural constraints: the background must be flat and non-textured [28], the illuminant vector must be known in advance [29], the illumination must be white [30-32], or the camera must be calibrated in a specific way [33]. The work of Barnard and Finlayson [34], discussed in detail in Chapter 5, only requires common illuminants – a very mild constraint indeed. They showed that for common illuminants, illumination changes across boundaries due to shadows exhibit certain properties. We use this idea as the starting point for our algorithm.

The discussion in this section shows that many image processing algorithms work well in laboratory settings. This thesis strives to compensate images for illumination without imposing any unnatural constraints (i.e., camera calibration, photometric stereo).

1.3 Thesis Contributions

This section summarizes the four main contributions of this thesis: two new illumination invariant representations called the $\hat{R}$ and $\hat{R}_{EDGE}$ images, an original method for detecting and removing specularities, and a novel shadow detection and removal technique.

The so-called illumination invariant $\hat{R}$ and $\hat{R}_{EDGE}$ images are based on the Retinex and can be used to improve the accuracy of computer vision applications such as face[1] and object recognition. As the $\hat{R}$ and $\hat{R}_{EDGE}$ images are remarkably shadow and specularity free, in this thesis they are used to locate material changes in an image.

A novel method for detecting and removing specularities is proposed. Based on an idea in [16], we detect *initial* areas that may be specularities, but very often either too little of the specularity is detected, or the specularity is detected along with the surrounding matte surface. This thesis introduces a technique whereby it is possible to grow a wavefront outwards from the center of the specularity to its boundary, or until a material boundary has been reached according to the $\hat{R}_{EDGE}$ image. Once the boundary of the specularity has been reached, the wavefront contracts inwards, coloring in as it contracts, until the specularity no longer exists.

A novel method for detecting and removing shadows is also proposed. For shadow detection, the theoretical basis for the color ratios across boundaries due to shadows is similar to but not the same as in [34], as our ratios allow shadows to be modeled in a manner more in keeping with physical laws. Furthermore, our extensive mathematical analysis of the properties of shadow boundaries and how they relate to color ratios is new. Moreover, in [34] an LUT is used that contains the ratios of *possible* shadow boundaries

[1] The $\hat{R}$ image has been used to improve face recognition accuracy by M. Gandhi at the Centre for Intelligent Machines, McGill University. Gandhi computed the $\hat{R}$ image for all faces in the Yale database, and then applied histogram fitting to ensure that all images had a similar dynamic range: 100% accuracy was obtained. This will be discussed in more detail in a forthcoming paper by Gandhi.

that can exist in nature. However, this thesis proposes a novel method that uses Support Vector Machines, whereby it is possible to identify *probable* shadow boundaries, not just possible ones. Furthermore, in this thesis it is shown that not only can shadows be extracted from these boundaries, but they can also be removed by coloring them with the average color of non-shadow neighbors of the same material.

In all cases – be it illumination invariance or the treatment of shadows and specualrities - no unnatural constraints are imposed. There is no need for more than one camera or a sequence of images, and no a priori information about the scene is necessary. The only condition that is imposed is that the illumination in the scene be from commonly existing indoor and outdoor illuminants.

1.4 Thesis Outline

Chapter 2 studies the effects of pre-processing images with the Retinex. The Retinex provides considerable dynamic range compression and color constancy, but it tends to gray out images [7-15]. The relationship between the Retinex and the image formation process is studied in depth, before discussing a scheme that restores color to images grayed out by the Retinex [9]. Color restoration was observed to weaken the color constancy of the original Retinex. Furthermore, it also had an arbitrary effect on pixel chromaticities, as also discovered in [6]. For example, pixels that were in the skin locus[2] [35-41] before applying the Retinex with color restoration were found to lie outside of it afterwards. It was decided that it would be safer to use the Luminance Retinex [6]. The Luminance Retinex only provides dynamic range compression while leaving the chromaticities of the original image unchanged. In order to deal with the issue of color constancy, it was necessary to restrict the illumination in scenes to common indoor and outdoor illuminants. Such illuminants have been found to lie in a well-defined region of chromaticity space known as the Planckian Locus [3,4,5], whose evolutionary basis is also discussed in this chapter. Finally, two new illumination invariant representions based on the Retinex, the $\hat{R}$ and $\hat{R}_{EDGE}$ images, are introduced.

In Chapter 3 specularity detection and removal is dealt with. The dichromatic reflection model is discussed, along with practical issues that limit its application. The work of Torres [16], which thresholds specularities using a binary mask in intensity-saturation space, is discussed and motivates our algorithm. Specularities are imperfectly detected with this method, but it is possible to determine their general location. It is shown that a specularity can be modeled in intensity space as a peaked mountain surrounded by a matte region that has the characterisitics of a flat surface such as a plain. From the peak of the specularity it is possible to grow a wavefront outwards until the bottom of the mountain has been reached or until a material boundary has been encountered as per the

[2] Recent research has shown [35-41] that the skin color distribution under common indoor and outdoor illuminants falls in a shell-shaped region in chromaticity space that is close to the Planckian locus. This shell-shaped region, where skin can be found, is often referred to as the skin locus [35-41]. Chapter 5 discusses the skin locus in more detail.

$\hat{R}_{EDGE}$ image. Once the wavefront expansion has stopped, the wavefront then grows inwards, coloring in as it grows, until the specularity is non-existent.

Chapter 4 studies in depth the mathematical properties of color ratios across boundaries due to shadows. It is shown that a Support Vector Machine can be used to identify probable shadow boundaries in an image. Problems are encountered with extremely strong shadows that are often severely clipped in color space, resulting in a loss or corruption of chromatic information. The mathematical properties of color ratios across boundaries between neighboring regions that were derived at the beginning of the chapter no longer hold in such cases. However, it is found that these problematical boundaries can be thresholded and an additional SVM can be used to differentiate between shadow and non-shadow borders among them. Once the probable shadow boundaries have been identified, this chapter describes how to extract shadows from the detected boundaries. Finally, the chapter concludes by demonstrating how the extracted shadows can be given the average color of neighbors that are of the same material.

Chapter 5 integrates the work done in the previous three chapters. The illumination compensation method - which consists of applying the Luminance Retinex, followed by specularity and shadow detection and removal – is tested on a wide variety of images with excellent results. The chapter then discusses the variety of applications that can benefit from illumination compensation. In particular, the benefits of illumination compensation to skin detection and face recognition are investigated. It is shown that because of shadows and specularities many skin pixels may go undetected. However, after compensating for illumination, these same pixels can be correctly identified as skin. It is also shown that the accuracy of face recognition can increase if images are first compensated for illumination using the method proposed in this thesis.

Chapter 2

Pre-Processing Images Using Retinex

2.1 Introduction

The Retinex [7-15] is an image enhancement algorithm that provides a high level of dynamic range compression[3] and color constancy. Moreover, it can be applied to arbitrary images without any prior knowledge of camera calibration. As a result, it has become a popular tool with which to pre-process images before applying further computer vision algorithms [42-44]. This chapter investigates the pros and cons of pre-processing images with the Retinex.

This chapter also describes two new illumination invariant representations based on the Retinex: the Retinex Uniformity Image ($\hat{R}$ image), and the $\hat{R}_{EDGE}$ image. The $\hat{R}$ image is a grayscale version of a given image that is free of the effects of varying illumination. In the $\hat{R}$ image, areas of uniform color are severely grayed out, but the edges between materials remain. As the graying occurs in areas of uniform color, the appellation Retinex Unifomity Image is coined. The $\hat{R}_{EDGE}$ image is an edge image that is derived from the $\hat{R}$ image. Figure 2.1 shows an example of both the $\hat{R}$ and the $\hat{R}_{EDGE}$ image. As both representations correct for illumination, they can be useful in many computer vision applications. In this thesis, the $\hat{R}_{EDGE}$ image is used to locate material changes in an image as part of the specularity detection process discussed later on in Chapter 3.

[3] When the dynamic range of a scene exceeds the dynamic range of the recording medium, the visibility of color and detail can sometimes be quite poor in the recorded image [9]. Dynamic range compression attempts to correct this situation by mapping a large input dynamic range to a relatively small output dynamic range [11].

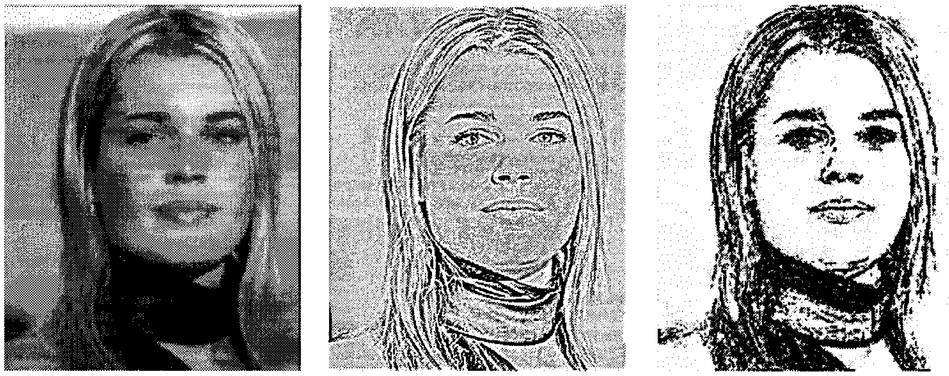


Figure 2.1 The $\hat{R}$ and $\hat{R}_{EDGE}$ images. **Left:** Original, **Middle:** $\hat{R}$ image, **Right:** $\hat{R}_{EDGE}$ image.

Chapter 2 is organized as follows: first Section 2.2 studies the image enhancement properties of the Retinex. The Retinex offers strong dynamic range compression and color constancy but the enhanced images tend to be grayed out. Section 2.3 investigates the relationship between the image formation process and the Retinex and Section 2.4 discusses how color can be restored to images that have been grayed by the Retinex. However, color restoration was observed to not only weaken the color constancy of the original Retinex, it also had an arbitrary effect on pixel chromaticities, as also found in [6]. It was decided that it would be safer to use the Luminance Retinex [6]. Section 2.5 describes the Luminance Retinex [6], which offers the dynamic range compression of the Retinex, but not the color constancy. Section 2.6 deals with the issue of color constancy as follows: instead of correcting for changes in illumination color, we can adapt to them by restricting the illumination in scenes to common indoor and outdoor illuminants, which tend to lie very close to a crescent shaped curve in x-y chromaticity space known as the Planckian Locus [3,4,5]. The evolutionary basis of the Planckian Locus is also discussed in this section. Finally, Section 2.7 introduces two new illumination-invariant representations: the $\hat{R}$ and $\hat{R}_{EDGE}$ images.

2.2 The Retinex

Many variants of the Retinex have been proposed over the years. The last version that Land proposed is now referred to as the Single Scale Retinex (SSR) [9-12] . The Single Scale Retinex for a point (x, y) in an image is defined [9-12] as being:

$$R_i(x,y) = \log I_i(x,y) - \log[F(x,y) \otimes I_i(x,y) \qquad (2.2.1)$$

where $R_i(x,y)$ is the Retinex output and $I_i(x,y)$ is the image distribution in the i th spectral band. In this thesis there are three spectral bands – one each for R, G and B. In the above equation the symbol $\otimes$ represents the convolution operator and $F(x,y)$ is the Gaus sian surround function:

$$F(x,y) = Ke^{-r^2/c} \qquad (2.2.2)$$

where $r^2 = x^2 + y^2$, and c is the Gaussian surround constant - analogous to the σ generally used to represent standard deviation. The Gaussian surround constant c is what is referred to as the scale of the Retinex. In previous research [14] it has been mathematically demonstrated that the Retinex algorithm provides color constancy by returning a ratio of the reflectances of a scene.

Figure 2.2 gives an example of the powerful color constancy that the Retinex offers, but also apparent are the benefits of dynamic range compression. Whereas the original image is rather dark, the enhanced image is not only less red, it is also reasonably brighter. The dynamic range compression that the Retinex offers is more evident in Figure 2.3. In both Figures 2.2 and 2.3 the poor lighting in the original images has been dramatically improved by applying the Retinex.

Figure 2.2 The Retinex corrects for illumination. **Left:** Original, **Right:** Result of applying Single Scale Retinex with surround *c=80*.

Figure 2.3 The Retinex offers strong dynamic range compression. **Left:** Original, **Right:** Result of applying Single Scale Retinex with *c=80*.

Typically, a small scale provides very good dynamic range compression, but at the cost of poorer color rendition, as graying is more common and pronounced in uniform zones of color, as these zones violate the gray world assumption upon which the Retinex is based [9]. Conversely, a large scale provides better color rendition, but at the cost of dynamic range compression [9], as seen in Figure 2.4.

Figure 2.4 The problem with the Retinex: a large scale provides better color rendition, but at the cost of dynamic range compression. **Left:** Original, **Middle:** SSR applied with small scale (c = 15), **Right:** SSR applied with large scale (c = 250)

The multiscale Retinex (MSR) tries to achieve a compromise between dynamic range compression and color rendition by combining the results of several scales. The multiscale Retinex output R_{MSR_i} for the i^{th} spectral channel is a weighted sum of N Single Scale Retinex outputs and is given in [9] as:

$$R_{MSR_i} = \sum_{n=1}^{N} \omega_n R_{n_i} \tag{2.2.3}$$

whereby R_{n_i} is the Single Scale Retinex output computed for the n^{th} scale c_n, and w_n is the weight associated with the n^{th} scale. In it is found that equally weighting the scales with one small (c_n = 15), one intermediate (c_n = 80), and one large scale (c_n = 250) is sufficient for most images. An example of applying the MSR is shown in Figure 2.5.

Figure 2.5 The Retinex generally results in desaturation of color. **Left:** Original, **Right:** Result of applying MSR

In Figure 2.5 the MSR enhanced image is clearly grayed out. In fact, all Retinex processing, whether SSR or MSR, generally results in desaturation of color to greater or lesser degrees, as seen in Figures 2.2-2.5. This graying effect occurs because in the MSR, a pixel's value in each channel is replaced with the ratio of its value to its neighbors. Thus, for pixels in areas where color is relatively uniform, the ratio in all three channels will be equal to one and look gray. Therefore, there is a need for a color restoration scheme. Color restoration is discussed in Section 2.4. Before dealing with color restoration, the relationship between the Retinex and the image formation process is first investigated in Section 2.3.

2.3 The Image Formation Process and the Retinex

This section investigates the relationship between the Retinex and the image formation process, as this relationship is a foundation for the mathematical analysis of shadows in Chapter 4. An image taken with a linear device such as a digital camera is composed of sensor responses whose value at a given pixel is given by [3] :

$$p_K = \int_{\lambda=400}^{\lambda=700} E(\lambda)S(\lambda)R_K(\lambda)d\lambda \qquad K = \mathrm{R,G,B} \tag{2.3.1}$$

where E is the illumination, S is the reflectance, and R_K is the camera sensitivity function. The camera sensitivity function can be assumed to be a Dirac delta function [3] with sensitivity at some wavelength, as follows:

$$R_K(\lambda) = \delta(\lambda - \lambda_K) \tag{2.3.2}$$

Finlayson [3] gives strong evidence that the Dirac assumption is valid over a wide range of sensors. The Dirac delta function has the well-known shifting property that gives:

$$p_K = E(\lambda_K)S(\lambda_K) \qquad K = \mathrm{R,G,B} \tag{2.3.3}$$

The reflectance component $S(\lambda_K)$ can be isolated if the illumination $E(\lambda_K)$ is known. Unfortunately, for arbitrary images the illumination is generally not known. Thus, recovering the original illumination and reflectance components in this fashion is akin to determining the two original factors given a product. If, however, the sensor response at a pixel is divided by its spatially weighted average value, the following is obtained [13,14]:

$$r_K = \frac{p_K}{\overline{p_K}} = \frac{E(\lambda_K)S(\lambda_K)}{\overline{E}(\lambda_K)\overline{S}(\lambda_K)} \qquad K = \mathrm{R,G,B} \tag{2.3.4}$$

where r_K is the new output pixel value. The bars denote the spatially weighted average value at a pixel, which is essentially the value that is obtained after Gaussian smoothing. For gradual changes in illumination the following holds true [14]:

$$E(\lambda_K) \approx \overline{E}(\lambda_K) \qquad K = \mathrm{R,G,B} \tag{2.3.5}$$

Consequently the output r_K is approximately a ratio of the reflectances of a pixel and its surroundings, thus providing independence from spectral variations in illumination [14] :

$$r_K \approx \frac{S(\lambda_K)}{\overline{S}(\lambda_K)} \qquad K = \mathrm{R,G,B} \tag{2.3.6}$$

In the literature it is often argued that the human visual system computes a ratio of an object's reflectance to the reflectance of its surround [7,8]. For many cases the above relation is an equality. For those cases where it is an approximation, the reflectance ratio dominates the spectral illumination variations [13,14]. Subsequently applying the logarithm to each pixel, as in Equation (2.3.7), has the effect of enlarging low intensity pixel values with respect to higher intensity pixel values, thereby further compressing the dynamic range. Thus:

$$r_K = \log \frac{p_K}{\overline{p}_K} \qquad K = \mathrm{R,G,B} \tag{2.3.7}$$

Recalling that the spatially weighted average value of a pixel is essentially the corresponding pixel value in a smoothed version of the image, the above can be re-written as follows:

$$r_K = \log(p_K) - \log[p_K \otimes F] \qquad K = \mathrm{R,G,B} \tag{2.3.8}$$

where $\otimes$ represents the convolution operator, and F is a Gaussian function. It is immediately apparent that the expression above is the last version of the Retinex that Land presented [9].

2.4 The Retinex: Problems with Color Restoration

In Section 2.2 it was found that Retinex processing, whether SSR or MSR, generally results in desaturation of color to greater or lesser degrees, as illustrated again in Figure 2.6. This graying effect occurs because in the MSR, a pixel's value in each channel is replaced with the ratio of its value to its neighbors. Therefore, for pixels in areas of uniform color the ratio in all three channels will be equal to one and look gray. Therefore, there is a need for a color restoration scheme.

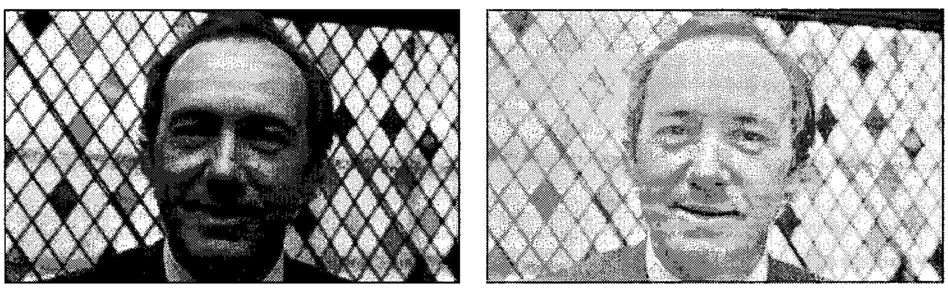


Figure 2.6 The Retinex generally results in desaturation of color. **Left:** Original, **Right:** Result of applying MSR

In [9] a color restoration function is proposed that successfully, in most cases, restores color to the desaturated images. A very similar function is proposed in [6]. An example of applying the MSR with color restoration (the MSRCR) is shown in Figure 2.7.

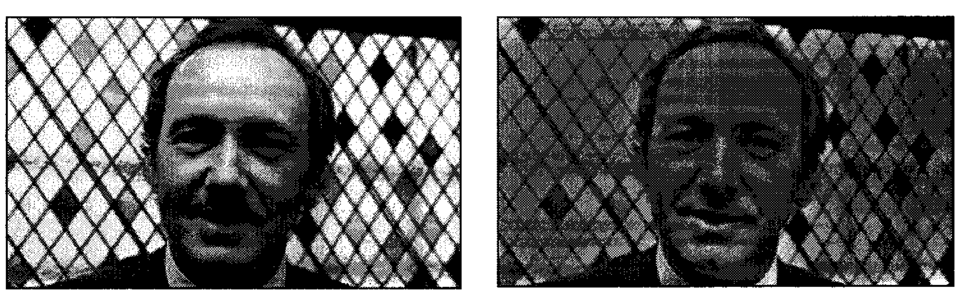


Figure 2.7 Color restoration successfully treats Retinex graying for most images. **Left:** Original, **Right:** Result of applying MSRCR with color restoration as per [6] .

The problem with these color restoration functions is that they use the chromaticities of the original image in order to restore color, which stands in direct contrast to the color constancy objectives of the Retinex. In fact, it was found that the stronger the color restoration, the weaker the color constancy. Even moderate amounts of color restoration significantly lessened the color constancy properties of the Retinex, as illustrated in Figure 2.8.

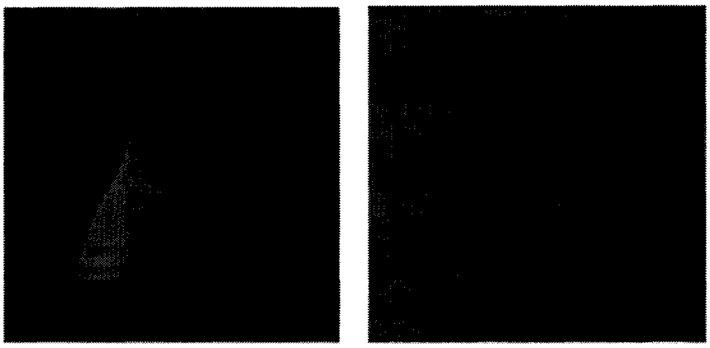


Figure 2.8 Color restoration weakens color constancy of the Retinex. **Left:** Original, **Right:** MSRCR

For most images, the dilution in color constancy is not very noticeable and is usually more than made up for by the gains in visual information and the great increase in color rendition [9]. The main reason that the dilution in color constancy is not very noticeable for most images is that most images are taken with standard illuminants. This will be discussed more in Section 2.5

Thus, the dilution of color constancy is of no great consequence if most images do not suffer as a result of it, and are still able to enjoy the benefits of dynamic range compression and contrast enhancement. Of greater concern is the following: the color restoration function changes image chromaticities in an unpredictable fashion [6]. Thus, it would be nice to get the dynamic range compression and contrast enhancement of the Retinex while at the same time having color fidelity (keeping the chromaticities of the original image). This is discussed in the next section.

2.5 Luminance Retinex

The fundamental problem with the Retinex, discussed in the previous section, can be summarized as follows: the stronger the dynamic range compression, the greater the resulting desaturation. Furthermore, any subsequent color restoration not only ends up weakening the original gains in color constancy, but also affects the image chromaticities in an unpredictable fashion. In this context, [6] proposes to separate the dynamic range component of the MSR from the color constancy component. The multiscale Retinex is only applied to the Luminance channel, thereby preserving the chromaticities of the original image while still providing dynamic range compression. Thus, the original formulation for the Single Scale Retinex of Equation (2.2.1) becomes:

$$R_L(x,y) = \log I_L(x,y) - \log[F(x,y) * I_L(x,y) \tag{2.5.1}$$

where L represents the intensity channel. The multiscale Luminance Retinex, hereafter referred to as the Luminance Retinex, is then simply the weighted sum of several luminance Retinex outputs using different scales. Identical to the original MSR [9] it uses three different scales with equal weights, and the scales remain unchanged (i.e., $c=15$, $c=80$, and $c=250$). As the Luminance Retinex is only applied to the intensity channel, the chromaticities of the original image remain unchanged. Figure 2.9 shows the result of applying the Luminance Retinex to an example image.

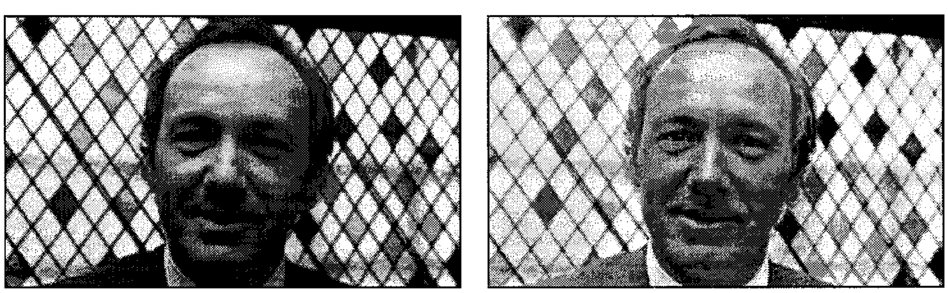


Figure 2.9 Luminance Retinex offers the dynamic range compression of the Retinex while preserving the original chromaticities of the image by only operating on the luminance channel. **Left:** Original, **Right:** Result of applying Luminance Retinex.

Thus, by using the Luminance Retinex, the dynamic range compression of the Retinex is obtained without encountering the various problems caused by color restoration.

2.6 Color Constancy, Evolutionary Psychology and the Planckian Locus

One of the great promises of the Retinex was the color constancy that it was supposed to provide. However, it has been seen that the color constancy that the original Retinex offers comes with some significant drawbacks. As a result, the decision was made to use the Luminance Retinex, which provides very strong dynamic range compression and contrast enhancement. But this begs the question: what ever shall be done about color constancy? In this section this question is studied in greater detail.

Color constancy, from the viewpoint of evolutionary psychology, is a very important adaptive function [45]. The ability to identify objects by their color across varying illumination conditions increases an organism's chances of survival. Evolutionary psychologist Shepard [46] proposes that characteristics of the world that have been present over the greatest amount of evolutionary time (e.g, sunlight) will be most deeply internalized. Thus, it stands to reason that the human visual system's color constancy will be optimal for the varying chromaticities of daylight [4,46] . Now, it is well established that the chromaticities of daylight vary during the day, depending on cloud cover and sun position, but they all fall very close to a crescent shaped curve in x-y chromaticity space known as the Planckian Blackbody Locus [3,4,5], as seen in Figure 2.10.

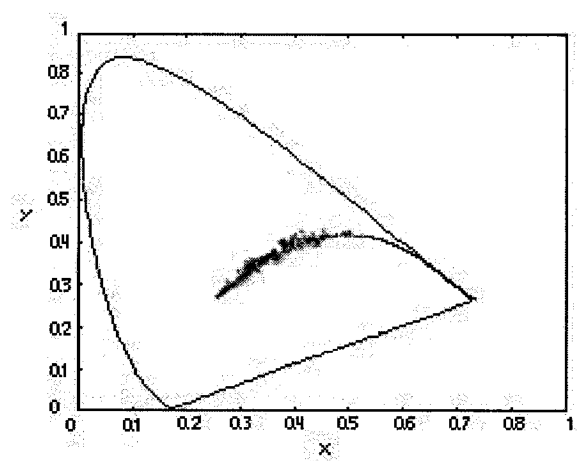


Figure 2.10 The Planckian Locus and common illuminants. Distribution of the x-y chromaticities of 172 common illuminants (in blue) and Planckian locus (in red), as per Finlayson [5]

The Planckian blackbody locus is produced by heating a blackbody radiator to various temperatures. What is of special interest here is that the Planckian locus includes not only the chromaticities of natural illuminants such as sunlight and blue skylight, but conventional illuminants from standard indoor yellow lighting (i.e., lightbulbs) to fluorescent lighting. In fact, Finlayson [5] plots 172 standard lights, as shown in Figure 2.10, including daylights and fluorescents, and finds that they cluster very tightly around the locus.

Thus, while it is possible to have a very saturated blue or red illuminant, in most practical situations the illuminant will fall on this locus. In fact, the lighting industry strives to manufacture illuminants with chromaticities that lie close to the locus [3] . Therefore, this thesis restricts itself to dealing with the vast majority of images, whose illuminants lie close to the locus.

2.7 The $\widehat{R}$ Image: A New Illumination-Invariant Image

This section proposes two new illumination-invariant representations based on the Retinex: the Retinex Uniformity Image ($\hat{R}$ image) and the $\hat{R}_{EDGE}$ image. The $\hat{R}$ image is a grayscale version of a given image that is free of the effects of varying illumination. In the $\hat{R}$ image areas of uniform color are severely grayed out, but the edges between different materials remain. As the graying occurs in areas of uniform color, the appellation Retinex Unifomity Image is coined for the $\hat{R}$ image. The $\hat{R}_{EDGE}$ image is an edge image that is derived from the $\hat{R}$ image. As both representations correct for illumination, they can be useful in many computer vision applications. In this thesis, the $\hat{R}_{EDGE}$ image is used in the specularity removal process discussed later on in Chapter 3.

The $\hat{R}$ image is based on the following property that the single scale Retinex exhibits: the smaller the scale, the greater the illumination invariance, and the greater the graying, especially in areas of uniform color. It stands to reason then, that if the scale is made extremely small the result should be an illumination invariant, albeit gray scale, image. Indeed, this is the case, as demonstrated in Figure 2.11. Moreover, since graying occurs in areas of uniform color, the unabbreviated appelation of this illumination invariant image is the Retinex Uniformity image.

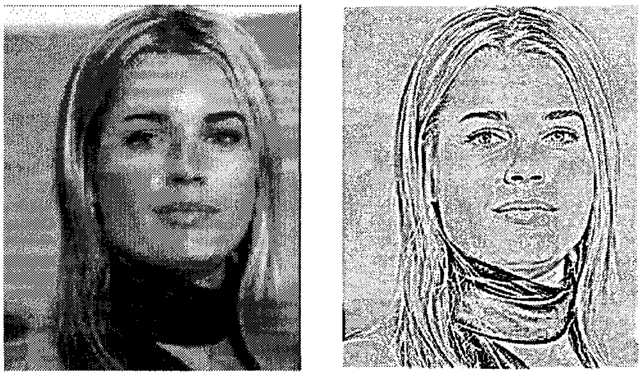


Figure 2.11 The $\hat{R}$ image. **Left:** Original **Right:** Result of applying Single Scale Retinex[4] with small scale ($c=2$)

To enhance the $\hat{R}$ image, an illumination-invariant edge image, the $\hat{R}_{EDGE}$ image, is obtained by thresholding the $\hat{R}$ image. As seen in Figure 2.12, the majority of pixels in the $\hat{R}$ image cluster around the peak of zero, and trail off very quickly on both sides, very much like a Laplacian distribution. The pixels clustering around the peak of the histogram represent pixels in areas of uniformity[5], which suggests that those pixels on either side of the peak must represent material (or very sharp shadow) edges[6] in the invariant image. It follows that material edges can be extracted from the $\hat{R}$ image by thresholding both sides of the peak. However, the $\hat{R}$ image is first gain-offset[7] corrected as per [13] so that the thresholding can be performed in the display domain[8]. Figure 2.12 shows histograms of both the raw Retinex and the gain-offset corrected Retinex output.

[4] After applying SSR with $c=2$ the image was converted to grayscale and contrast stretched for display purposes.
[5] As the Retinex averages a pixel's value with that of its surround, pixels in areas of uniform color will have a value of one. However, since the Retinex also subsequently takes the logarithm, they will in fact have a value of zero.
[6] It was found that the vast majority of shadows were successfully removed by computing the $\hat{R}$ image. However, very sharp shadow edges could not always be fully removed by the $\hat{R}$ image.
[7] Gain-offset correction as per [13] maps an input dynamic range [a,b] to [0,255].
[8] The display domain is [0,255], which includes all possible RGB or grayscale pixel values that a digital image can have.

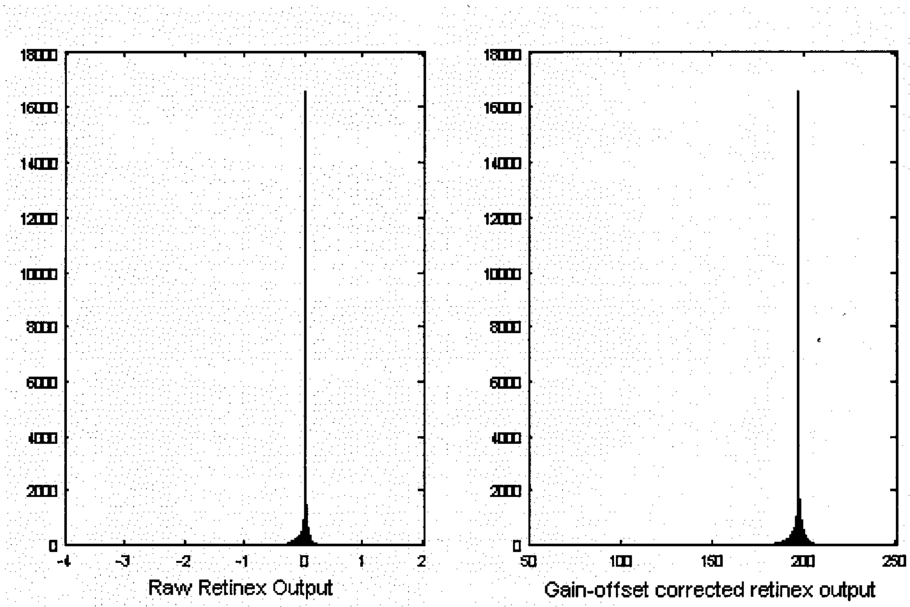


Figure 2.12 Retinex Intensity histograms. **Left:** Raw Retinex output, **Right:** Gain-offset corrected Retinex output

Before formally defining the $\widehat{R}_{EDGE}$ image, a few definitions are presented. First T is defined as the gray level value that corresponds to the peak of the histogram. Next, uniform areas are posited to exist between the thresholds t_1 and t_2, which are related to the peak T as follows :

$$\begin{aligned} t_1 &= T - k \\ t_2 &= T + k \end{aligned} \tag{2.7.1}$$

where k is a very small integer. The transformation between $\widehat{R}$ and $\widehat{R}_{EDGE}$ can then be formally defined as follows:

$$\begin{aligned} &if \;\; \widehat{R}(x,y) > t_1 \cap \widehat{R}(x,y) < t_2 \Rightarrow \widehat{R}_{EDGE}(x,y) = 255 \\ &if \;\; \widehat{R}(x,y) < t_1 \cup \widehat{R}(x,y) > t_2 \Rightarrow \widehat{R}_{EDGE}(x,y) = 0 \end{aligned} \tag{2.7.2}$$

In Equation (2.72), pixels whose grayscale values lie between t_1 and t_2 (i.e., within uniform areas) are arbitrarily made white in the $\widehat{R}_{EDGE}$ image, while everything else is deemed a material edge and made black. When k is too large, many valid material edges are mistakenly identified as being uniform areas. Similarly, when k is too small, too many false edges are detected and the image becomes rather cluttered. It was found that $k = 3$ was a good compromise, as shown in Figure 2.13.

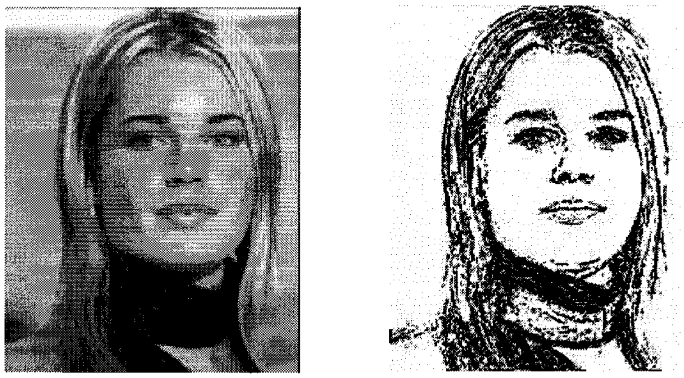


Figure 2.13 The $\hat{R}_{EDGE}$ image. **Left:** Original, **Right:** $\hat{R}_{EDGE}$ Image

Thus, two new and powerful illumination invariant representations have been discovered: the $\hat{R}$ image, and the $\hat{R}_{EDGE}$ image. As the $\hat{R}_{EDGE}$ image indicates where in the image material boundaries lie, any particular location in a given image can be cross-referenced with its corresponding $\hat{R}_{EDGE}$ image in order to determine whether or not the location corresponds to a material boundary. This property of the $\hat{R}_{EDGE}$ is used in Chapter 3 to detect and remove specularities.

2.8 Conclusions

In this chapter it was found that while the Retinex provided very strong dynamic range compression, it also produced color constancy at the price of excessive graying. While, the color restoration function was suitable for image enhancement purposes [9] , its drawback was that it effectively undermined the original goal of color constancy and also changed image chromaticities in an unpredictable manner [6]. In light of these discoveries, instead of correcting changes in illumination color, this thesis chose to adapt to them by restricting itself to lights that lie close to the Planckian Locus. After all, most standard illuminants lie close to the Planckian locus [3,4,5], and the lighting industry strives to manufacture illuminants close to this locus [3]. Furthermore, a decision was made to pre-process images with a variant of the Retinex: the Luminance Retinex [6], which provides powerful dynamic range compression. The Luminance Retinex enhances an image by applying the Retinex to the intensity channel, thereby offering dynamic range compression, while preserving the chromaticities of the original image.

The investigation outlined above led to the discovery of two illumination invariant representations based on the Retinex: the $\widehat{R}$ image, and the $\widehat{R}_{EDGE}$ image. The $\widehat{R}$ image is remarkably free of illumination effects, as is the $\widehat{R}_{EDGE}$ image, which is an edge image derived from the $\widehat{R}$ image. Both representations can be used in conjunction with other computer vision applications. Also, as the $\widehat{R}_{EDGE}$ image indicates where in the image material boundaries lie, any particular location in a given image can be cross-referenced with its corresponding $\widehat{R}_{EDGE}$ image in order to determine whether or not the location corresponds to a material boundary. In Chapter 3 this property of the $\widehat{R}_{EDGE}$ image is used to detect and remove specularities.

Chapter 3

Detecting and Removing Specularities

3.1 Introduction

The classical approach to dealing with specularities in images begins with the Dichromatic Reflection Model [17] for inhomogeneous dielectrics. According to this model, for a given surface the RGB color signal C at a pixel is a linear combination of the light C_I reflected at the material *surface* (C_I is interchangeably referred to as *highlight*, *specularity* or *interface* reflection) and the light C_B reflected from the material *body* (C_B is referred to as *body* or *matte* reflection):

$$C(x,y) = m_I(\theta) \cdot C_I(x,y) + m_B(\theta) \cdot C_B(x,y) \tag{1}$$

where m_I and m_B are the corresponding weight factors which depend on the geometry of the scene, including the angle of incidence of the illuminant and the viewing angle.

Several techniques have been proposed to separate pixels into their dichromatic components [17,18] and produce a so-called *intrinsic* matte image with the specularity component removed. The problem with this approach is that for the highlight removal process to work well, only materials which show the same reflection properties can be included [19]. Otherwise, the color clusters of the different materials will overlap in RGB space. Thus, a prior segmentation of the image is required.

Statistical color image segmentation methods such as [47] and [48] do not account for the image formation process. As a result, the segmentation fails to identify highlights as belonging to a given material. Physics-based color segmentation methods, on the other hand, try to take the image formation process into account. In [17,20] the RGB cube is searched for skewed-T-shaped clusters in an effort to determine the number of materials in a scene, segment them, and then separate them into their respective matte and interface components. As the analysis of 3-dimensional color space is costly, in [19] the complexity of the task is reduced to the scrutiny of a 2D *u-v* chromaticity space. Similarly, in [21-23] clusters are detected in HSI space based on peaks and valleys in histograms. All of these methods are successful when the objects in a scene have very saturated colors (such as plastic spheres and cups) and exist in a laboratory setting where the background is uniform, usually black. The use of photometric stereo [19,24-27] along with the dichromatic reflection model, however, shows some promise in being able to successfully analyze and remove highlights in images of complex scenes. As this thesis is concerned with removing the highlights in a single image, photometric stereo techniques do not concern us here

This chapter describes a novel method for detecting and removing specularities in images. Based on an idea in [16], initial areas that may be specularities are detected. However, quite often either too little of the specularity is detected or it is detected along with the surrounding matte surface. The technique introduced in this chapter describes how a wavefront can be grown outwards from the center of the specularity to its boundary, or until a material boundary has been reached according to the $\widehat{R}_{EDGE}$ image. Upon reaching the boundary of the specularity, the wavefront contracts inwards. As it contracts, it colors in the specularity until it no longer exists. The method successfully removes specularities from typical images as seen in Figure 3.1.

Figure 3.1 Removing specularities. **Left**: Original, **Right**: After specualrities have been removed using the method in this thesis.

This chapter is organized as follows: Section 3.2 discusses how potential specularities can be detected in an image. Section 3.3 studies the relationship between specularities and their matte surroundings. Section 3.4 describes how a wavefront can be grown outward from the center of a specularity to its boundary, and Section 3.5 shows how to remove the specularity by coloring the wavefront inwards.

3.2 Detecting Potential Specularities

A method for the detection of specularities in color images has been proposed recently, whereby certain relationships between intensity and saturation are exploited [16]. The method constructs a bi-dimensional histogram of an image called the *MS* diagram, where *M* represents intensity and *S* stands for saturation. In [16] it is found that highlights are located in a well defined region of *MS* space, independent of hue. In this thesis, intensity is referred to as *I*, and the *MS* diagram will henceforth be alluded to as the *IS* diagram or *IS*-space. The authors of [16] then construct a binary mask in *IS*-space to detect potential specularities. This method is used with some modifications in order to locate seed points in the general vicinity of specularities.

We now proceed to describe the method in detail. In this thesis, the definition of intensity is the same as in [16]:

$$I = \frac{1}{3}(R + G + B) \tag{4}$$

For saturation, the following well-known expression for saturation is utilized:

$$S = \frac{Max(R,G,B) - Min(R,G,B)}{Max(R,G,B)} \tag{5}$$

Before creating the *IS* diagram, the luminance channels of the images are processed by histogram equalization in [16] to ensure that the upper limit of the dynamic range is 255[9]. The Luminance Retinex [6] is applied in place of histogram equalization. Figure 3.2 shows some typical *IS* diagrams.

[9] In this thesis, the upper limit is set to 1.

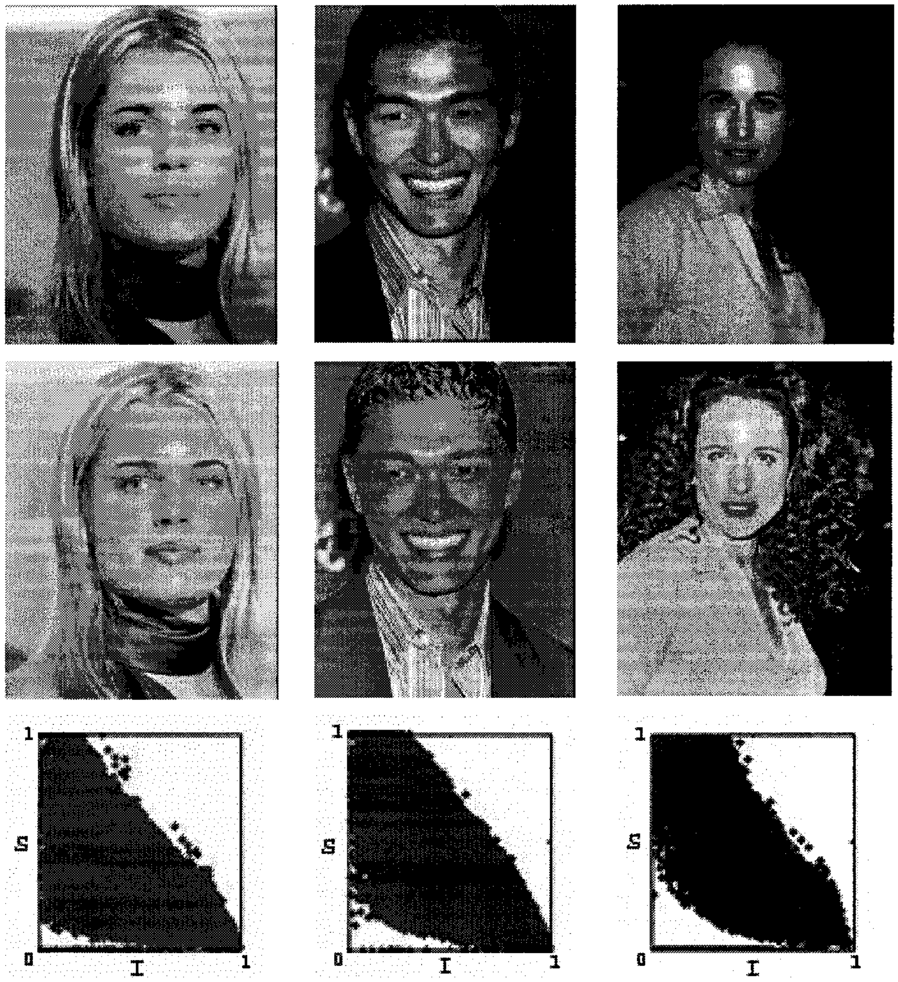


Figure 3.2 The IS Diagram. **Top:** Original, **Middle:** Luminance Retinex output, **Bottom:** IS Diagram of Luminance Retinex output

As specularities tend to be bright and desaturated, they cluster in the bottom right-hand corner of the *IS* plane. By analyzing a representative selection of images, the authors of [16] produce a binary mask (shown in Figure 3.3) in IS-space that can be used to segment highlights.

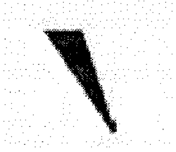


Figure 3.3 Binary mask used in [16].

Unfortunately, it has not been possible to replicate their results, as the boundaries of the mask are not altogether apparent in the paper. Instead, a new mask was created by inspecting a set of 40 images from the Internet that have specularities of varying degrees. Specularities were manually identified in these images and the peak intensity of each specularity was plotted in an *IS* binary mask (Figure 3.4). In other words, the most intense point within each specularity was plotted in the bi-variate histogram.

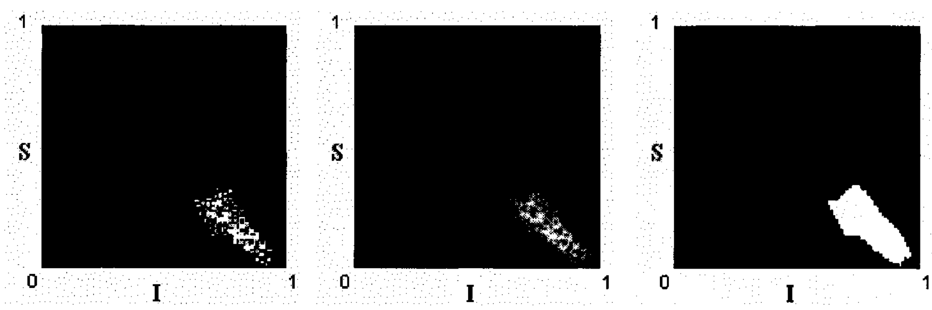


Figure 3.4 Creating the Binary Mask, **Left:** Seed points, **Middle:** Lines (red) fitted[10] to the extremities of the seed points, **Right:** Final Mask

The reason that the most intense point of each specularity is plotted in the bi-variate histogram is as follows: since our plan is to grow outward from detected seed points to the boundaries of the specularities, then at a minimum the peak (point of highest intensity) of each specularity should be detected. Now, the mask created with the seed points (left-most in Figure 3.4) has many holes in it. Therefore, in order to increase our chances of detecting specular areas, the holes were filled in by fitting lines to the extremities of the cluster and then using the enclosed area as the mask (Figure 3.4). Potential specularities were then thresholded with the binary mask (Figure 3.5).

[10] The equations of the lines take the form $S=mI+b$ where $m_1=1.117$, $m_2=0.875$, $m_3=0.48$, $m_4=-0.74$, $m_5=-1.24$, $m_6=2.22$, $m_7=0.01$, $b_1=-1.05$, $b_2=-0.28$, $b_3=0.48$, $b_4=0.69$, $b_5=1.22$, $b_6=1.69$, $b_7=0.65$

Figure 3.5 Detecting specularities using the IS binary mask. **Top:** Original, **Bottom:** Specularities thresholded using the IS mask.

Figure 3.5 shows that this approach successfully locates specular regions in an image. However, it becomes immediately apparent that specularities are often confused with non-specular materials that are also bright and desaturated. Secondly, the approach often fails to successfully identify the limits of the specularity boundary. As a result, either the full specularity fails to be detected (under-detection) or the bounds of the specularity are exceeded (over-detection). These issues are dealt with in the next section.

3.3 The Mountain and the Plain: The Relationship Between Specularities and their Matte Surroundings

In Section 3.2, it was shown that specularities can be detected using a binary mask in MS space. While the detection is not perfect because of under- and over-detection, in most cases the center of the specularity (a seed point) has been correctly detected. Therefore, it is proposed that the center of the specularity be used as an initial condition for an expanding wavefront, one that grows outwards in all directions with a constant velocity. A series of advancing wavefronts that ultimately take the shape of a container (as in Figure 3.6) are envisioned. The boundaries of the container will either be the boundary of the specularity or a material boundary.

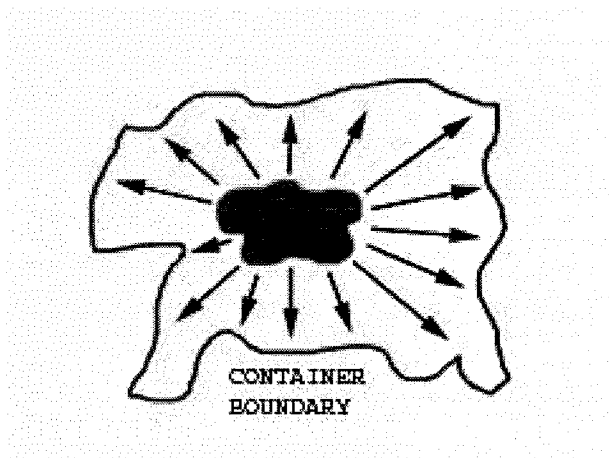


Figure 3.6 An expanding wavefront that takes the shape of a container.

Thus, the wavefront has two stopping conditions: either it reaches the specularity boundary or it encounters a material boundary. The reasoning behind this is as follows: the specularity detection scheme in Section 3.2 can confuse specularities with bright, desaturated materials. Thus, if the detected seed point mistakenly lies on such a material and is not actually a specularity, the wavefront is allowed to expand at most to the boundary of the material, thereby preventing the expansion from continuing indefinitely. Also, even if the seed point actually lies on specularity, stopping the expansion at a material boundary is a safety net to catch and terminate an expansion that has gone awry and advanced past the specularity boundary. In this case, the wavefront can be prevented from expanding indefinitely by stopping the expansion at a material boundary.

Of the two stopping conditions (specularity boundary and material boundary), the latter is not an issue, as the $\widehat{R}_{EDGE}$ image can simply be cross-referenced in order to determine whether or not a material boundary has been reached. Of greater concern is finding a way to determine when the boundary of the specularity has been reached.

The properties of hundreds of specularities from a wide variety of images were examined and it was found that within specular regions, the different color variable systems (e.g., RGB, rg, HSI) tend to vary quite erratically. However, the surrounding matte regions tend to have a more stable color. It was also found that specularity can be modeled as a 3D surface such as a mountain. Figure 3.7 shows typical *intensity* plots of some specularities. Since specularities are brighter than their surroundings, they form "mountains" in intensity space and the surrounding "plain" corresponds to the matte surface or stable color, for which we are searching. The point where the mountain ends and the plain begins is the specularity boundary.

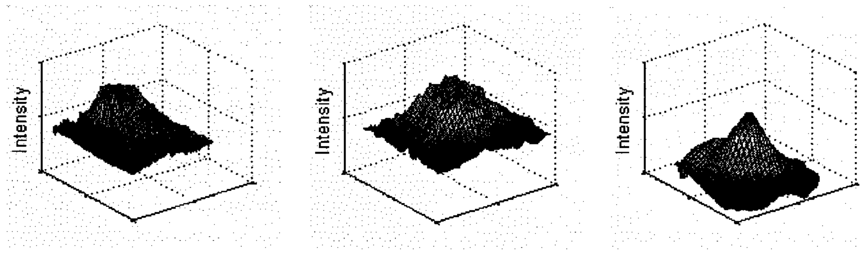


Figure 3.7 A specularity can be thought of as a 3D surface such as a mountain. Above, intensity plots of three typical specularities.

The goal then is to propagate every point on the wavefront outward until it either hits a stable color or hits an $\widehat{R}_{EDGE}$ image boundary. Furthermore, a stable color has been reached when the plain surrounding the specularity mountain is encountered.

3.4 Growing Outwards to the Specularity Boundary

In the previous section it was shown that specularity can be modeled in intensity space as a mountain surrounded by a planar matte region. The aim is to find a way to descend this mountain, stopping once a stable color (i.e., the plain) has been reached. Now, since the specularity is essentially a topographic surface, for a given specularity it would be helpful to examine a contour map of the intensity levels, such as the one shown in Figure 3.8.

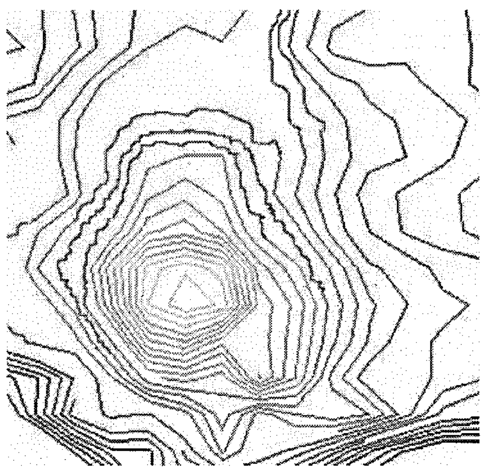


Figure 3.8 Contour map of specularity. Each contour level can be thought of as a wavefront.

In Figure 3.8 it is seen that each contour level can be viewed as a wavefront at a given stage in the propagation. Thus the contour map depicts a series of advancing wavefronts that expand outward from the peak of the specularity. As the mountain is descended from its peak to lower and lower contour levels, the total size of the region will increase slowly. However, upon reaching the plain, the region size will increase much more quickly. In Figure 3.9 it is seen that the plots of contour level versus total region size are typically parabolic. Furthermore, the slope of the curves begins to sharply increase at or very close to the bottom of the mountain. In order to detect this sharp increase in slope a 3-point quadratic is fitted to the data at every iteration using the method of Least Squares and the slope is extracted. Since the line of best fit considers the last three points, it gives a good approximation of the direction or slope of the data at any given iteration[11]. A sharp increase is defined arbitrarily as being 50% or more, and the wavefront is grown outward from the peak of the specularity mountain to successively lower contour levels by using a classical recursive flood-fill algorithm [53]. Figure 3.10 shows an example of growing wavefronts outward from the peaks of potential specularities.

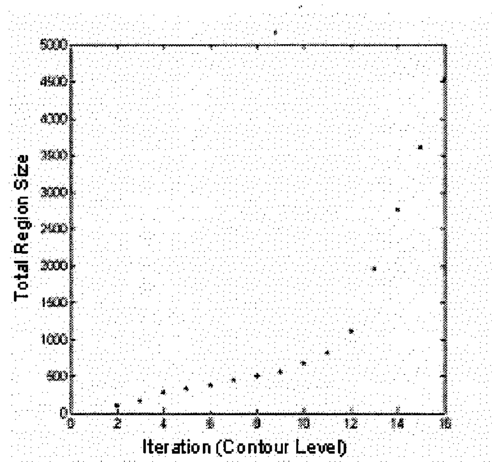


Figure 3.9 Plot of contour level versus total region size for some typical specularities. Each contour level has an intensity of 1% less than the previous one. The sharp increase in slope marks the point where the plain(matte) region begins and the mountain(specularity) ends.

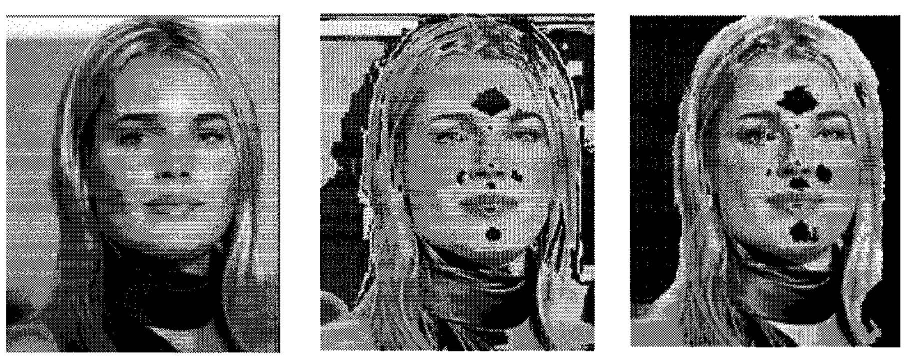


Figure 3.10 Results of growing wavefronts outward from the peaks of potential specularities. **Left**: Original image, **Middle**: Potential specular areas detected with IS mask. **Right**: After growing outward from the peak of each potential specularity.

[11] Note that for Iteration 1 the slope is undefined and for Iteration 2 it is not examined.

An issue encountered while growing outward is that sometimes the wavefront expansion is cut short too early, as shown in Figure 3.11 where a specularity seed point on the forehead is detected, but the wavefront nonetheless fails to expand to the boundary of the specularity. In fact, the wavefront barely seems to expand at all. The reason this occurs is that it was assumed that the total region size increases in a relatively linear fashion until the sharp change in slope at the specularity boundary. This initial linear increase is referred to as the "specularity line". Similarly, the term "matte line" refers to the linear increase after the sharp change in slope. The graphs in Figure 3.9 fit this model quite well. However, the curves of the specularities whose expansions are cut short indicate that a certain number of specularities can be modeled as having a leading cluster of points with relatively low or zero slope. These points are designated as noise, as illustrated in Figure 3.12.

Figure 3.11 Sometimes the expansion is cut short too early. **Left**: Original, **Middle**: Specular areas detected with IS mask. A seed point that is barely visible is detected on the forehead. **Right**: After growing outwards: the wavefront fails to expand from the seed point to the boundary of the specularity.

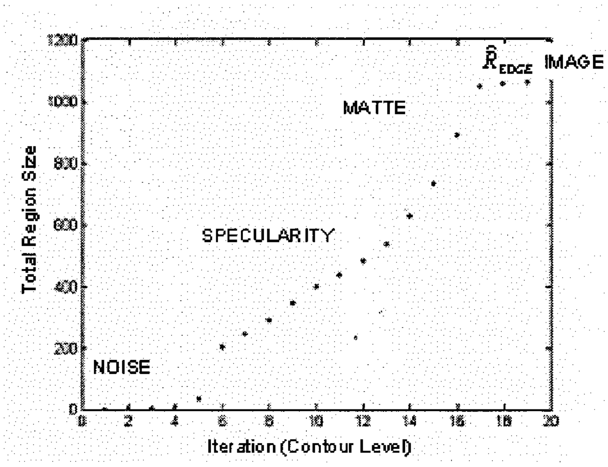


Figure 3.12 A more complete model: contour level vs. total specularity region size. Very often, during the first few iterations the region size increases very minimally. Compare with Figure 3.9

In Figure 3.12, it is only after this initial zero slope, or noise line, that the specularity line appears, followed by the matte and $\hat{R}_{EDGE}$ image lines[12]. Ignoring this initial horizontal slope yields vastly superior results, as illustrated in Figure 3.13.

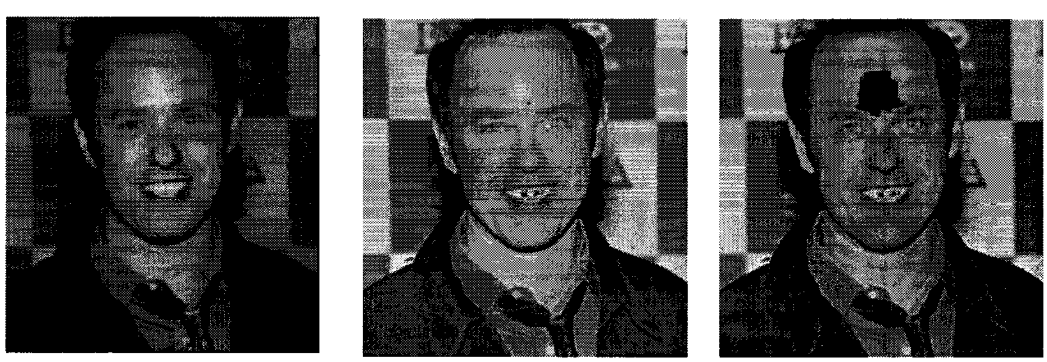


Figure 3.13 Clipping the noise line prevents the expansion from being cut short. **Left**: Original, **Middle**: Growing in and out. **Right**: Growing in and out after first clipping the noise line.

The results of detecting specularities with some other images are shown in Figure 3.14.

[12] At a certain point the wavefront expansion must ultimately stop as it cannot expand past material $\hat{R}_{EDGE}$ image boundaries. As a result the total region size will remain constant, resulting in a plateau or line of zero slope. This line is the $\hat{R}_{EDGE}$ image line.

Figure 3.14 Detecting specularities by growing wavefronts outward from their peaks. **Top:** Original, **Bottom:** Detected specular regions.

Now that specular regions in an image are successfully identified, the next task is to color in these regions with the color at their boundaries. This is discussed in the next section.

3.5 Coloring Inwards

All wavefronts are colored inwards by repeatedly finding the new interior boundary of the detected regions and coloring each boundary pixel with the average color immediately outside the boundary. The algorithm, illustrated in Figure 3.15, can be stated as follows:

1. Find the boundary of the specularity wavefront.
2. Color the specularity boundary by coloring each specularity boundary pixel with the average color of neighboring pixels outside the specularity wavefront.
3. Repeat 1 and 2 until the specularity no longer exists.

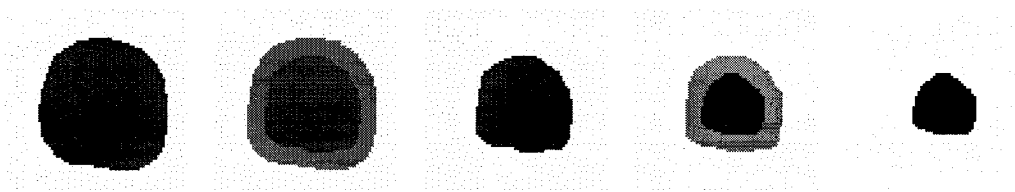


Figure 3.15 Coloring a specularity inwards. From left to right: the specularity becomes smaller and smaller as the wavefront boundary is repeatedly colored inwards. The detected specular region is red, and the wavefront boundary is green.

Also, when part of the specularity boundary coincides with an $\hat{R}_{EDGE}$ image boundary (i.e., a material boundary), it is not colored inwards for the following reason: coloring a specularity inwards from a material boundary can result in the specularity being colored in incorrectly. The reason for this can be seen in Figure 3.16: when part of a specularity boundary coincides with a material boundary, the surrounding matte region has not been reached for that portion of the specularity. Thus, that portion of the specularity should not be colored inwards as it will incorrectly be colored inwards with the color of specularity.

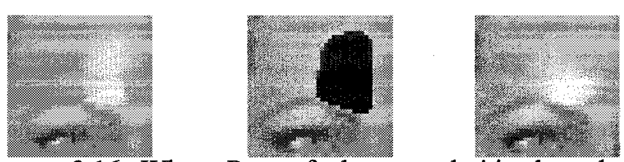


Figure 3.16 When Part of the specularitiy boundary coincides with a material boundary. In this example the specularity shares a boundary with the eyebrow. **Left**: Original, **Middle**: Specularity in red, **Right**: After coloring inwards.

Specularities were colored inwards and it was found that due to noise in the $\widehat{R}_{EDGE}$ image, small "holes" in the original image were sometimes left uncolored, resulting in a Swiss cheese-like result, as shown in Figure 3.17.

Figure 3.17 $\widehat{R}_{EDGE}$ image noise sometimes results in artifacts. **Top:** Original, **Bottom:** After removing specularities. Note the small "holes" on the forehead that have been left uncolored.

The solution is to fill the holes before coloring in by eliminating tiny (size of 4 pixels or less) $\widehat{R}_{EDGE}$ image clusters[13] that lie within a detected area. Rather than being valid material boundaries these tiny clusters of pixels in the $\widehat{R}_{EDGE}$ image tend to be noise.

Figure 3.18 shows the results of applying the specularity removal algorithm presented in this chapter to a wide variety of images. Not only has pre-processing with the Luminance

[13] A cluster of pixels is defined as a group of interconnected pixels.

Retinex brightened the images, but also the largest and most severe specularities have been removed. As skin detection schemes often fail to correctly identify highlights, any segmentation scheme applied to the processed images should now be more efficacious.

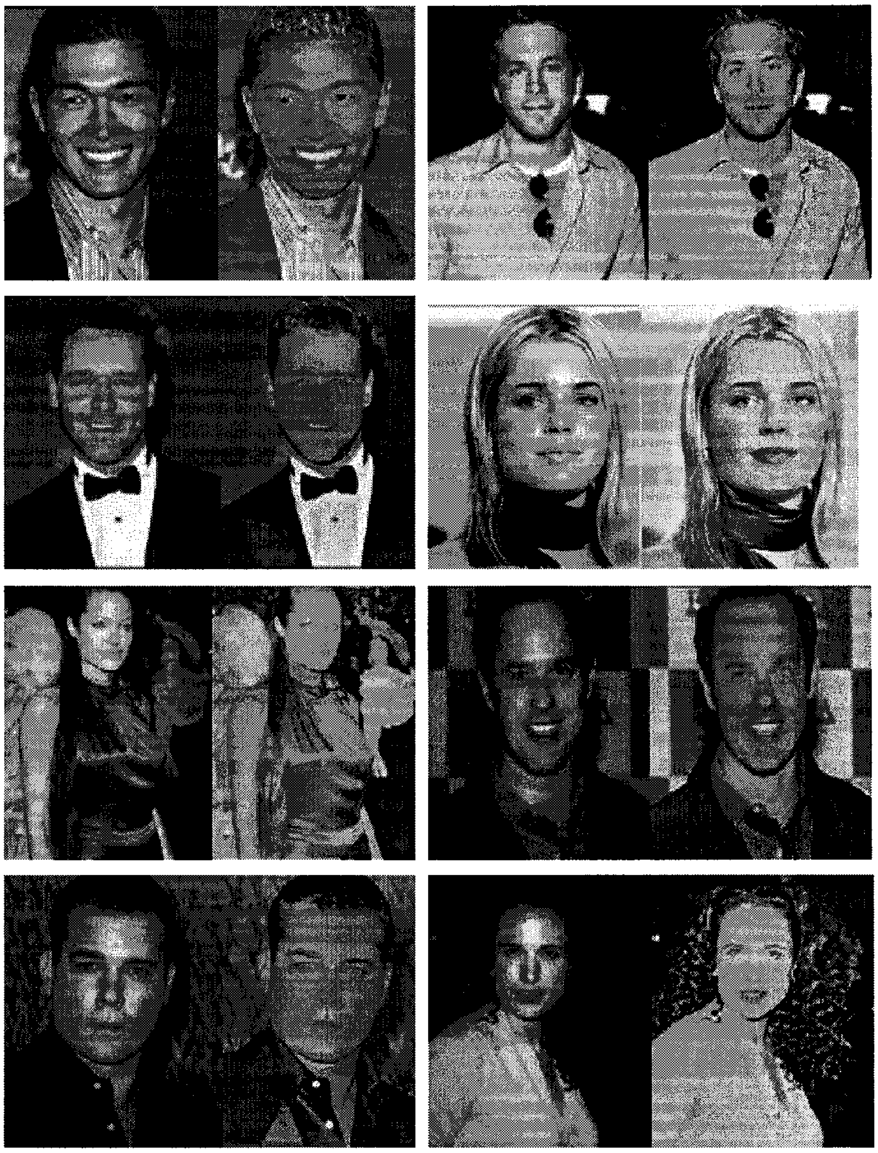


Figure 3.18 Specularity detection and removal. For each pair of images, on the left is the original image, while on the right is the result of applying the Luminance Retinex [6] and then detecting and removing specularities.

3.6 Conclusions

The Introduction in Section 3.1 pointed out that the vast majority of methods specifically targeting specularities are either based on the dichromatic model or use photometric stereo. As also discussed in Section 3.1, methods based on the Dichromatic Model have been shown in the literature to work well, but have been tested only on laboratory images. Since photometric stereo requires the use of several images it was not investigated further since a single image method was sought for this research.

The novel specularity detection and removal technique proposed in this thesis does not require more than a single camera or a sequence of images. Neither is a prior segmentation of the image required. Also, the technique is not limited to "toy images" in laboratory settings - it is robust and can process complex scenes. Furthermore, this chapter also presents two new illumination invariant representations: the $\widehat{R}$ and $\widehat{R}_{EDGE}$ images. The specularity detection and removal approach consists of finding the seed regions of specularities using saturation and intensity, from which we grow outwards either to the boundary of the specularity or until a material boundary is encountered in the newly discovered $\widehat{R}_{EDGE}$ image. Once the boundary of the specularity has been reached, the waveform moves inwards, coloring in as the region contracts. The process of coloring continues until the specularity no longer exists. Any subsequent skin detection scheme should show more success, as specularities will no longer be present to confound the segmentation.

Chapter 4

Detecting and Removing Shadows

4.1 Introduction

Natural images often contain shadows and these usually confound their analysis by current computer vision approaches. It has been found to be difficult to distinguish between shadow and non-shadow regions when they have similar hue, saturation and intensity. In [28] the authors propose a method to identify and classify shadows as being cast or self. Unfortunately, two unnatural constraints are imposed: first, the background must be flat and non-textured; second, there must be no occlusions between shadows and objects. Another shadow detection method [29] requires an advance knowledge of the illuminant vector. Several illumination invariant color spaces have been proposed and used for shadow detection [30-32], but they all have the constraint of requiring white illumination. One way to avoid this constraint is to white-balance the camera[49,50], but our approach does not require any camera calibration, thereby making it more practical for applications.

The work of Barnard and Finlayson [34] does not require camera calibration. They showed that shadow boundaries have certain interesting properties. Specifically, illumination changes across shadow boundaries were shown to exhibit color ratios that were different from the ratios across material boundaries. In [34] a lookup table was used to keep track of *possible* illumination changes across shadow boundaries. However, the method in this thesis automatically identifies *probable* illumination changes, not just possible ones. Our work is based on the research in [34]. We, however, use Support Vector Machines to identify *probable* shadow boundaries in typical images; shadowed regions are inferred from this boundary information. The shadowed regions are then removed by assigning them the color of non-shadow neighbors of the same material. The method successfully removes shadows from typical images as seen in Figure 4.1.

Figure 4.1 Removing shadows. **Left**: Original, **Right**: After shadows have been removed using the method in this thesis. Note that the highlights still remain.

This chapter is organized as follows: first Section 4.2 studies the properties of color ratios across boundaries between regions in a segmented image. Then Section 4.3 focuses on the relationship between these color ratios and boundaries caused by shadows. Section 4.4 describes how a Support Vector Machine can be used to differentiate shadow boundaries from those due to a change in material. Section 4.5 discusses training a Support Vector Machine to identify shadow boundaries in a segmented image. Some boundaries are incorrectly classified, and Section 4.6 deals with these problem boundaries. Section 4.7 shows how shadows can be extracted from their corresponding boundaries, and Section 4.8 describes how to remove them by assigning them the color of neighboring pixels of the same material.

4.2 Color Ratios Across Region Boundaries

Consider an image segmented into N regions, $R_1 \ldots R_N$, where B_{ij} is the boundary between neighbors R_i and R_j, as shown in Figure 4.2.

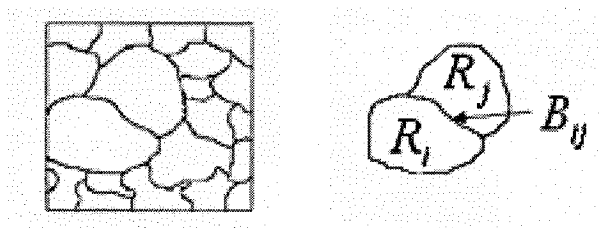


Figure 4.2 Image regions and boundaries. **Left:** Image segmented into N regions $R_1 \ldots R_N$, **Right:** The boundary between neighbors R_i and R_j is B_{ij}.

Equations (2.3.1) to (2.3.3) in Section 2.3 mathematically describe the relationships between the image formation process, illumination and reflectance. Equation 2.3.3 is repeated below as Equation (4.2.1), as it is the foundation for the mathematics that follows in this chapter:

$$p_K = E(\lambda_K) S(\lambda_K) \qquad K = R, G, B \qquad (4.2.1)$$

In the above equation, p_K is the sensor response at a given pixel, $S(\lambda_K)$ is the reflectance and $E(\lambda_K)$ is the illumination. Next, consider Figure 4.3 where p_{Ki} is a pixel in R_i adjacent to B_{ij}. In other words, it lies on one side of the boundary B_{ij} between R_i and R_j, specifically the side that belongs to R_i. Similarly, let p_{Kj} be a pixel in R_j and adjacent to B_{ij}.

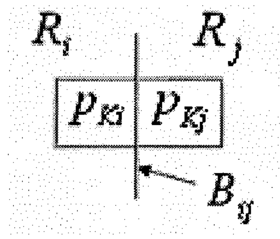


Figure 4.3 Pixel p_{Ki} is in R_i and adjacent to B_{ij},

If the sensor response at p_{Ki} is divided by the sensor response at p_{Kj}, the following ratio is obtained:

$$ratio_{Kij} = \frac{p_{Ki}}{p_{Kj}} = \frac{E_i(\lambda_K)S_i(\lambda_K)}{E_j(\lambda_K)S_j(\lambda_K)} \qquad K = R, G, B \qquad (4.2.2)$$

This thesis focuses on the detection of shadow boundaries after an image has been segmented into regions. In this case, if both pixels lie on the same material, one property they have in common is their reflectance. Since $S_i(\lambda_K) = S_j(\lambda_K)$, then:

$$ratio_{Kij} = \frac{p_{Ki}}{p_{Kj}} = \frac{E_i(\lambda_K)}{E_j(\lambda_K)} \qquad K = R, G, B \qquad (4.2.3)$$

which is a ratio of the illumination intensities in each channel, independent of reflectance. If it is then assumed that the spectral distribution of the illumination is the same for both pixels, and that it is only the intensity of the illumination that changes, then:

$$E_i(\lambda_K) = c_{ij}E_j(\lambda_K) \qquad c_{ij} \geq 0 \qquad K = R, G, B \qquad (4.2.4)$$

where c_{ij} depends on the geometry of the scene, such as the angle of incidence of the illuminant and the viewing angle. The ratio can then be rewritten as:

$$ratio_{Kij} = \frac{p_{Ki}}{p_{Kj}} = \frac{c_{ij} E_j(\lambda_K)}{E_j(\lambda_K)} = c_{ij} \qquad c_{ij} \geq 0 \qquad K = R, G, B \qquad (4.2.5)$$

The above equation states that, given two pixels located on both sides of a boundary between two neighboring regions with the same surface reflectance, and illuminated by the same spectral distribution, the ratios of the two pixels will be the same in all three color channels.

In practice it is rare that Equation (4.2.5) holds true. In natural images two pixels p_{Ki} and p_{Kj} with $S_i(\lambda_K) = S_j(\lambda_K)$ will differ in hue, saturation and chromaticity as well as intensity. The reason that Equations (4.2.4) and (4.2.5) indicate that p_{Ki} and p_{Kj} only differ in intensity is because ambient illumination has not been accounted for. The illumination E in Equations (2.3.1) to (2.3.3) and also in Equation (4.2.1), actually consists of an ambient component and a contribution from the light source. Thus:

$$E(\lambda_K) = E_A(\lambda_K) + E_L(\lambda_K) \qquad K = R, G, B \qquad (4.2.6)$$

In Equation (4.2.6), E_A signifies the ambient illumination and E_L represents the illumination from the light source, both of which have different spectral distributions. As the intensity and spectral distribution of the ambient illumination are often modeled as being identical everywhere in the scene [51-53], the constant A can be substituted for $E_A(\lambda_K)$. The ratio in Equation (4.2.5) then becomes:

$$ratio_{Kij} = \frac{p_{Ki}}{p_{Kj}} = \frac{c_{ij} E_{Lj}(\lambda_K) + A}{E_{Lj}(\lambda_K) + A} \qquad c_{ij} \geq 0 \qquad K = R, G, B \qquad (4.2.7)$$

The ratio in Equation (4.2.7) is explored further in the next section.

4.3 Shadows Across Region Boundaries

Equation (4.2.7) in the previous section gives the ratio between pixels p_{Ki} and p_{Kj} located on either side of a boundary B_{ij} between two neighboring regions R_i and R_j. Furthermore, both regions have the same surface reflectance and are illuminated by the same light source, E_L and ambient light, A. Consider the case when p_{Ki} is in shadow and p_{Kj} is not, as in Figure 4.4.

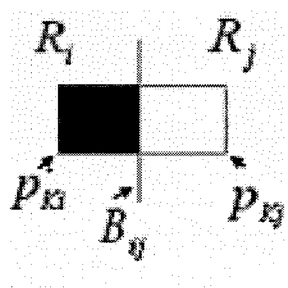


Figure 4.4 R_i and R_j have the same surface reflectance.
Pixel p_{Ki} is in shadow and pixel p_{Kj} is not.

Substituting p_{KS} for p_{Ki} and p_{KN} for p_{Kj} results in the following expression:

$$ratio_{KSN} = \frac{p_{KS}}{p_{KN}} = \frac{c_{ij}E_{LN}(\lambda_K) + A}{E_{LN}(\lambda_K) + A} \qquad 0 \le c_{ij} < 1 \qquad K = R, G, B \qquad (4.3.1)$$

where now $c_{ij} < 1$ since the pixel that is in shadow must have a lower intensity than the pixel not in shadow. Equation (4.3.1) gives the ratio of two pixels across a boundary between two neighboring regions with the *same* surface reflectance, where one region is in shadow and the other is not.

There is one very common occurrence, however, that Equation (4.3.1) fails to model: a region in shadow that is a neighbor to another region (of the same material) that is in shadow. Figure 5 illustrates how this could be possible. When an object casts a shadow, part of it will be in umbra, while another part will be in penumbra. The umbra is that region of the shadow where the shadowing object blocks all of the light from the source, whereas the penumbra is that region adjoining the umbra where the shadow is only partial. Therefore, it is actually possible for a region in umbra to be adjacent to a region (of the same material) in penumbra. Furthermore, Figure 4.5 also illustrates that the penumbra does not have a constant intensity: it gradually increases in intensity from shadow to light. In an image that has been segmented into regions, a penumbra may be segmented into several adjacent regions. Thus, it is also possible for two regions (of the same material) in penumbra to be adjacent to each other.

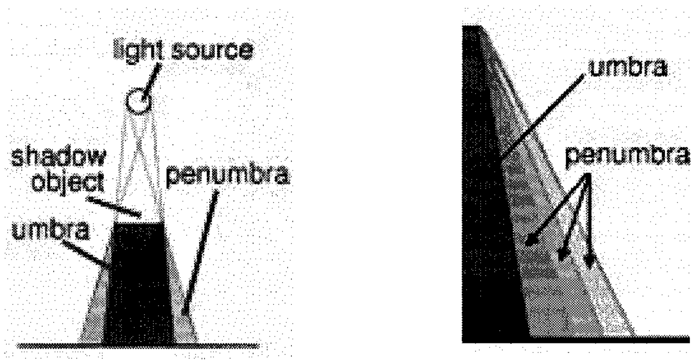


Figure 4.5 Umbra and penumbra. **Left:** A non-point light source will produce three distinct lighting areas [54] in a scene: directly lit, partially lit (penumbra), and not lit at all (umbra). **Right:** The intensity of the penumbra gradually increases from shadow to light.

If both neighboring regions are in shadow, Equation (4.2.7) reduces to the following:

$$ratio_{KSS1} = \frac{p_{KS1}}{p_{KS2}} = \frac{c_{ij}E_{LS2}(\lambda_K) + A}{E_{LS2}(\lambda_K) + A} \qquad 0 \le c_{ij} < 1 \qquad K = R, G, B \qquad (4.3.2)$$

In Equation (4.3.2), p_{KS1} and p_{KS2} are pixels on either side of the boundary between neighboring regions R_1 and R_2 that are both in shadow, as shown in Figure 4.6.

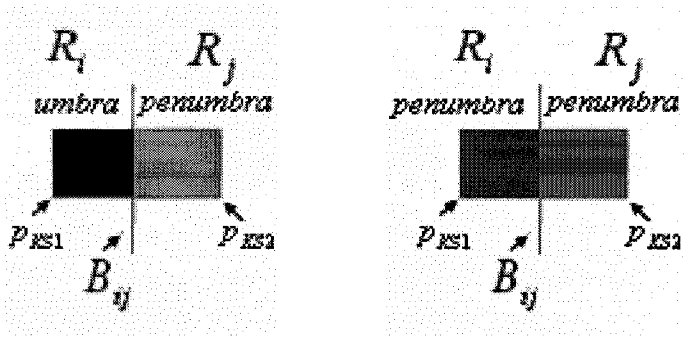
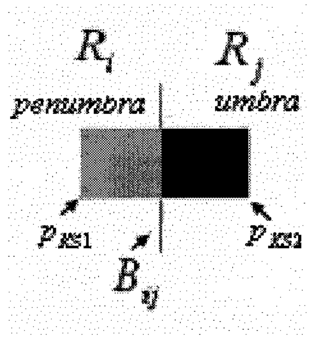


Figure 4.6 In Equation (4.3.2) regions R_i and R_j have the same surface reflectance and pixel p_{KS2} is brighter than pixel p_{KS1}. **Left:** Pixel p_{KS1} is in umbra and pixel p_{KS2} is in penumbra. **Right:** Both pixels are in penumbra.

When $c_{ij} = 0$ then p_{KS1} is in umbra and p_{KS2} is in penumbra. When $0 < c_{ij} < 1$ then both p_{KS1} and p_{KS2} are in penumbra. However, Equation (4.3.2) does not permit p_{KS2} to be in umbra. Moreover, in Equation (4.3.2) p_{KS2} is always brighter than p_{KS1} because a greater fraction of the light source reaches it. Therefore, Equation (4.3.3) is introduced to model the complementary case in which p_{KS2} may be in umbra and where p_{KS1} is brighter than p_{KS2}:

$$ratio_{KSS2} = \frac{p_{KS1}}{p_{KS2}} = \frac{E_{LS1}(\lambda_K) + A}{c_{ij}E_{LS1}(\lambda_K) + A} \qquad 0 \le c_{ij} < 1 \qquad K = R, G, B \qquad (4.3.3)$$

Equation (4.3.3) not only allows p_{KS2} to be in umbra, it also allows p_{KS1} to be brighter than p_{KS2}, as shown in Figure 4.7.



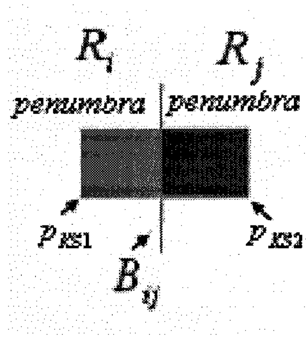


Figure 4.7 In Equation (4.3.3) regions R_i and R_j have the same surface reflectance and pixel p_{KS1} is brighter than pixel p_{KS2}
Left: Pixel p_{KS2} is in umbra and pixel p_{KS1} is in penumbra.
Right: Both pixels are in penumbra.

The three ratios expressed in Equations (4.3.1), (4.3.2), and (4.3.3) are the theoretical foundations of the shadow identification scheme proposed in this thesis. The next section discusses how these ratios can be used with a Support Vector Machine [55-58] to identify shadow boundaries[14].

[14] A shadow boundary is a boundary between two neighboring regions of the same material, whereby at least one region is in shadow. Thus, a shadow boundary in this thesis may be between a shadow region and a non-shadow region as in Equation (4.3.1). It may also be between two shadow regions, as in Equations (4.3.2) and (4.3.3).

4.4 Shadow Boundaries and Support Vector Machines

In [34] a shadow identification method is proposed that uses a color ratio similar to but not exactly the same as Equation (4.3.1). The derivation in [34] uses the term "shadow illumination" in place of ambient illumination, and shadows are assumed to be illuminated exclusively by ambient light. When converted to the nomenclature used in this thesis, the color ratio in [34] can be expressed as follows:

$$ratio_{KSN'} = \frac{p_{KS'}}{p_{KN'}} = \frac{A}{E_{LN}(\lambda_K) + A} \qquad K = R, G, B \qquad (4.4.1)$$

where $ratio_{KSN'}$ is the ratio between pixels $p_{KS'}$ and $p_{KN'}$ which lie on either side of a boundary between a shadow and a non-shadow region, A is the ambient illumination, and E_{LN} is the contribution of the light source. As seen in Equation (4.4.1), the expression derived in [34] models a shadow pixel as being in umbra: the pixel is not affected by the light source - it is illuminated only by ambient light.

In [34] the authors restrict themselves to common indoor and outdoor illuminants that, they show, form a cone in RGB space. All light sources and ambient illuminants are assumed to lie within this cone. Thus, all possible ratios between two pixels across a boundary between neighboring regions belonging to the same material, where one pixel is in shadow and the other is not, can be pre-computed and stored in a lookup table [8]. Each ratio in the LUT represents a possible change in illumination due to a shadow. Subsequently, a given image is segmented into regions and the color ratios for all boundaries between neighboring regions are examined. If a color ratio for a boundary appears in the LUT, then it is *possible* that the boundary is the result of a shadow. However, many further tests need to be performed to determine to what degree of confidence the boundary can be safely considered an *actual* shadow boundary [34]. The authors suggest that it is difficult to determine with any single test whether the boundary is really a shadow:

> "To estimate the plausibility that an edge is a shadow edge we use a number of tests, each of which has a score associated with it ... We are currently working on a more principled scoring, but we note that preliminary results indicate that the exact numbers are not that important. We remind the reader that the final score for the boundary is the maximal score found among all tests"

One difficulty with the reasoning in [34] is that the pixel on the shadow side of the boundary is modeled as being in umbra. Thus, ratios across boundaries between soft shadows (i.e., penumbra) and non-shadow regions of the same material are not in the LUT. Furthermore, ratios across boundaries between neighboring shadow regions of the same material are not in the LUT either. We address this issue by using Equations (4.3.1), (4.3.2) and (4.3.3) as the theoretical basis for color ratios across shadow boundaries. These equations permit us to model shadow regions as both umbra and penumbra.

A greater problem that arises in [34] is that the color ratios in the LUT indicate *possible* illumination changes due to shadows, but the color ratios by themselves give no evidence of *probable* illumination changes due to shadows. A ratio in the LUT can also be due to a change in material [34]. In order to better comprehend this, we examine how a shadow illumination change[15] across a boundary between neighboring regions can have the same ratio as a material change. We define a new ratio– the ratio between two pixels, p_{K1} and p_{K2} on either side of a boundary between two regions belonging to different materials:

$$ratio_{K12} = \frac{p_{K1}}{p_{K2}} = \frac{S_1(\lambda_K)[E_{L1}(\lambda_K)+A]}{S_2(\lambda_K)[E_{L2}(\lambda_K)+A]} \qquad K = R,G,B \qquad (4.4.2)$$

Since the spectral distribution of the light source is the same for both pixels, it is only the intensity of the light source that changes. Thus:

$$E_{L2}(\lambda_K)=cE_{L1}(\lambda_K),\ \ c>0 \qquad K = R,G,B \qquad (4.4.3)$$

[15] A shadow illumination change refers to a change in illumination due to a shadow on a given material.

Substituting Equation (4.4.3) into Equation (4.4.2) gives:

$$ratio_{K12} = \frac{p_{K1}}{p_{K2}} = \frac{S_1(\lambda_K)[E_{L1}(\lambda_K) + A]}{S_2(\lambda_K)[c_{12}E_{L1}(\lambda_K) + A]} \qquad K = R, G, B \qquad (4.4.4)$$

In Equation (4.4.1) the color ratio in [8] was converted to the nomenclature used in this thesis. A comparison of Equation (4.4.1) and Equation (4.4.4) shows that values of $S(\lambda_K), E_L(\lambda_K), A$, and c can be chosen such that the two ratios can indeed be made the same. Even after restricting $E_L(\lambda_K)$ and A to common illuminants, combinations of $S(\lambda_K)$ and c can result in identical ratios for shadow ilumination changes (Equation 4.4.1) across a boundary and material changes (Equation 4.4.4) across a boundary. The same holds true for Equations (4.3.1), (4.3.2) and (4.3.3), which are the color ratios derived in this thesis that correspond to shadow boundaries. Thus the method in [34] leads to ratios that represent possible shadow boundaries that may also represent material boundaries[16].

Instead of using an LUT, in this thesis we train a Support Vector Machine [55-58] using color ratios to identify which boundaries between neighboring regions are shadow boundaries. An LUT will include many ratios of shadow boundaries that are unlikely to occur alongside more common and probable ones. An SVM, on the other hand, can be trained with color ratios to accurately identify *probable* shadow boundaries in a principled manner.

[16] A material boundary is a boundary between two neighboring regions of different material, as in Equations (4.4.2) and (4.4.4).

4.5 Training A Support Vector Machine To Identify Shadow Boundaries

This section describes how an SVM can be trained to differentiate shadow from non-shadow boundaries. To accomplish this, a wide variety of images containing shadows was collected from the Internet and separated into training and test sets. Next, the Luminance Retinex [6] was applied to all of the images. These were then segmented into regions using EDISON (Edge Detection and Image Segmentation), a mean-shift color image segmentation program [59-61]. Figure 4.8 shows the results of applying EDISON to a sample image.

Figure 4.8 The EDISON program for color image segmentation. **Left:** Original, **Middle:** Luminance MSR image, **Right:** EDISON applied to Luminance MSR image.

Next, shadow and non-shadow regions in the training set were manually identified, and boundaries in the images were classified as belonging to one of the four categories detailed in Table 1.

1	Same Material	Shadow / Non-shadow
2	Different Materials	Shadow / Different Material
3	Same Material	Shadow / Shadow
4	Everything else	Don't care

Table 1 Training the SVM: each shadow boundary is classfied as belonging to one of four categories.

From the training set, a training file consisting of the features of thousands of boundaries was created. The SVM was given the following features for each boundary: the three ratios[17] across the boundary (one for each color channel), r and g chromaticity, and intensity. It was found that the number of entries in each class must be roughly the same in order to obtain good results. Also, non-shadow boundaries must include a large number of borderline cases in order for the SVM to accurately separate the data. Therefore, the boundaries of many materials that look like shadows, such as eyebrows and hair in faces, were included[18]. The training file was then submitted to the SVM, which learned the difference between the various boundary types[19]. More specifically, the SVM constructed a set of hyperplanes that separated the data in the feature hyperspace[20]. Finally, using the hyperplanes it had created, the SVM was used to classify boundaries in images from the test set as being either shadow or non-shadow, as shown in Figure 4.9.

One limitation of the approach described so far is that for extremely strong shadows, severe clipping at the lower end of the spectrum results in the loss or corruption of chromatic information. The equations in Sections 4.2, 4.3 and 4.4 no longer hold in these cases, as is apparent in the last row of Figure 4.9. In the next section, these problematical boundaries are identified and an additional SVM is then used to determine whether or not they are truly shadow boundaries.

---

[17] These are found in Equations (4.3.1, 4.3.2, 4.3.3).

[18] The final training file consisted of 13658 boundaries from 50 images. The specific frequencies of the various border types were as follows:

Type 1: (Shadow/Non-shadow): 2915
Type 2 : (Shadow/Different Material): 3230
Type 3 : (Shadow/Shadow): 2938
Type 4 : (Everything Else): 4574

[19] While training the SVM, the linear kernel was investigated along with the different types of nonlinear kernels. Specifically, experiments were conducted with the polynomial, sigmoid, and radial basis functions. The best results were achieved using the radial basis function (RBF) with a cost, C, of 64 and a gamma, γ, of 64. It was not known beforehand which C and γ were optimal. Consequently, a parameter search was done using the "grid search" utility that comes with the libsvm package, in which pairs of (C, γ) are tried and the one with the best cross-validation accuracy is picked. It was also found that the results were slightly worse if the training file was scaled with the svm-scale utility provided by libsvm. This may be because the absolute values of the ratios have some inherent meaning. In any case, before writing to the training file, intensity was manually scaled to lie between 0 and 1. Furthermore, r and g chromaticity is constrained to be between 0 and 1 by definition. For the ratios, both the raw values and normalized values were given. Therefore the total number of features was nine: the three ratios across the boundary (raw *and* normalized), r and g chromaticity, and intensity.

[20] The total number of support vectors, 3999 out of a possible 13658, indicated that over-fitting was not a problem, which was corroborated by a 5 fold cross-validation accuracy of 92.55%.

Figure 4.9 Results of the shadow boundary detection method. **Left:** Original, **Middle:** Luminance Retinex [6] image after segmentation, **Right:** Boundaries detected on segmented image. Shadow/Shadow boundaries are blue, while all other shadow boundaries are green.

4.6 Treating Problem Boundaries

It was found that the vast majority of the mistakenly identified boundaries tended to have low intensity and were located close to the achromatic axis in RGB color space. This is seen in the histograms in Figure 4.10. In this section it is shown that problem boundary histograms can be thresholded and an additional SVM can be used to differentiate between shadow and non-shadow borders.

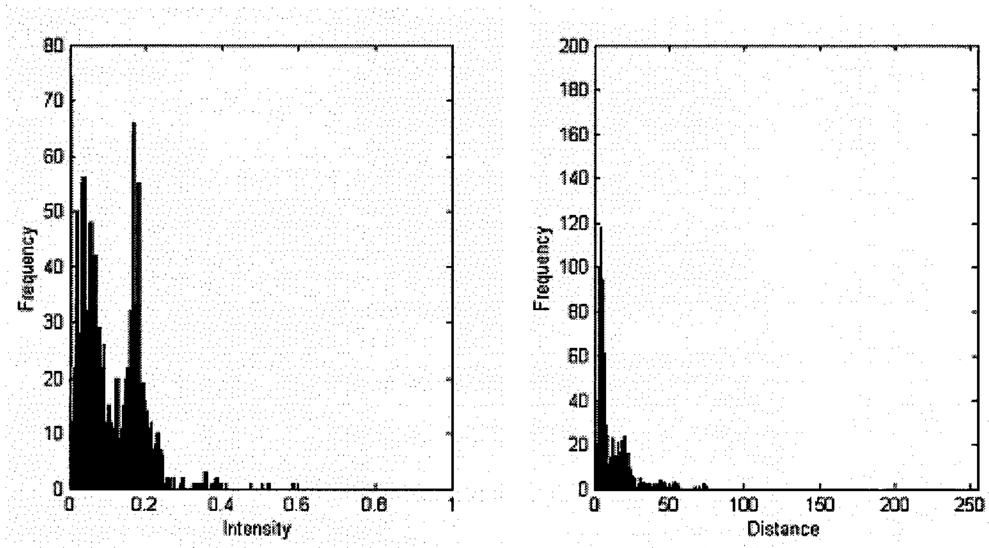


Figure 4.10 Histograms of problem boundaries. **Left:** Intensity, **Right:** Distance from achromatic axis.

In order to determine the histogram thresholds in Figure 4.10, ROC (Receiver Operating Characteristic) curves were plotted and EER (Equal Error Rate) values for each threshold were found.[21] The training set consisted of 13658 boundaries of which 1017 were problem boundaries. FAR and FRR values for many different thresholds were computed and plotted for both intensity and distance from the achromatic axis, as shown in Figure 4.11.

[21] The following definitions are useful:
FAR: False Accept Rate. For a given threshold, the probability that a regular boundary is falsely identified as a problem boundary.
FRR: False Reject Rate. For a given threshold, the probability that a problem boundary is rejected, and thus falsely identified as a regular boundary.
EER: Equal Error Rate. The threshold at which FAR=FRR.

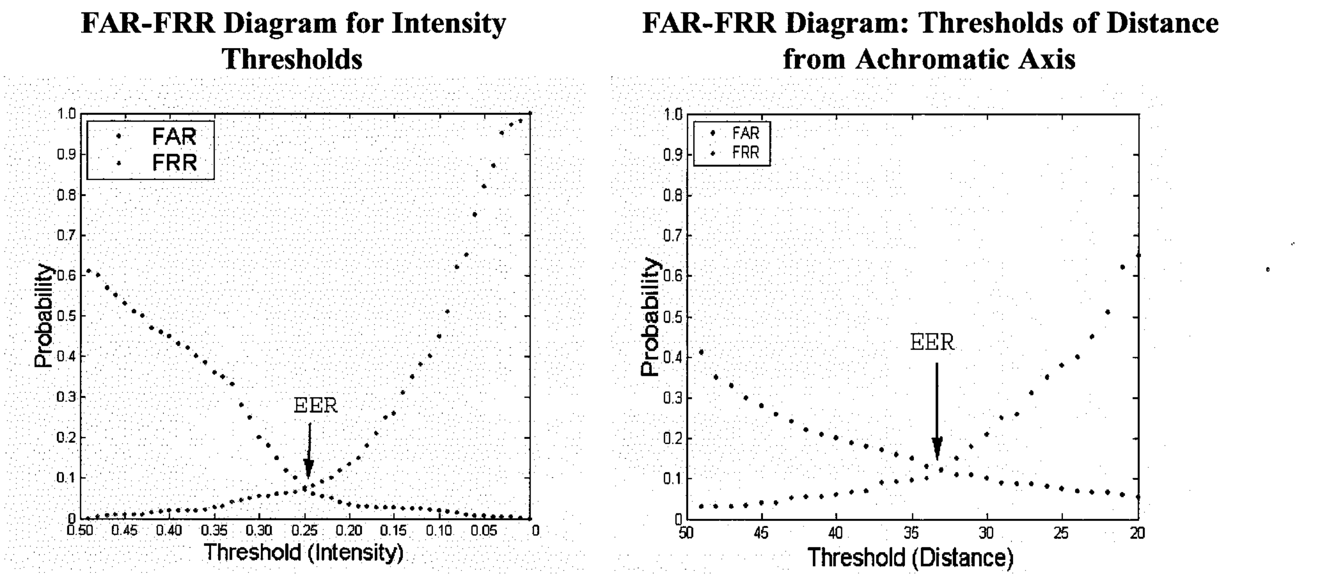


Figure 4.11 FAR-FRR Diagrams. **Left:** FAR-FRR diagram for intensity thresholds. **Right:** FAR-FRR diagram for thresholds of distance from the achromatic axis.

In Figure 4.11 the EER for intensity is 7% and occurs when the threshold is 0.26, or 26%. Similarly, the EER for distance from the achromatic axis is 12%, corresponding to a distance of 33 pixels. A more common, way of representing the above information is to construct an ROC curve as done in Figure 4.12.

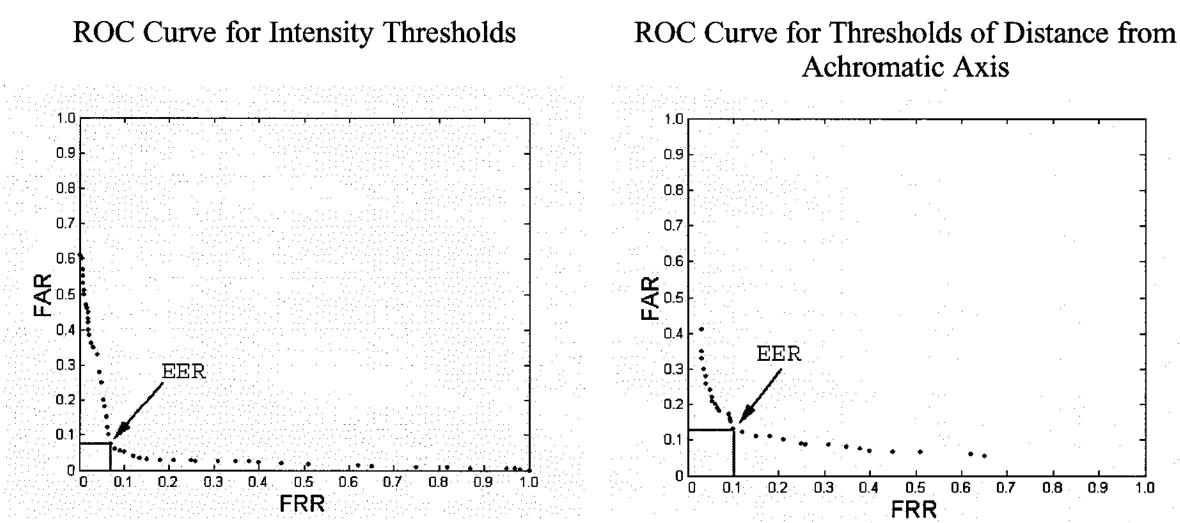


Figure 4.12 ROC curves. **Left:** Intensity thresholds. **Right:** Thresholds of distance from the achromatic axis.

Using these thresholds, the problem boundaries within the original training set were determined and an additional SVM was trained to differentiate between shadow and non-shadow borders[22]. As illustrated in Figure 4.13, treating problem boundaries separately made it possible to more accurately determine whether or not they were truly shadow boundaries.

Figure 4.13 Treating problem boundaries separately. **Left:** EDISON of Luminance MSR image, **Middle:** Shadow boundaries detected with the original SVM, **Right:** Shadow boundaries detected after thresholding the problem boundaries and treating them separately.

[22] For each misidentified boundary, the SVM was provided the following information: intensity, distance from the achromatic axis, r and g chromaticity, and whether or not it was a shadow boundary. If it was a shadow boundary, it was indicated what category it fell in, using the categories defined in Table 1. The training file consisted of the 1017 misidentified borders. The specific frequencies of the various border types were as follows:

Type 1 (Shadow/Non-shadow, same material): 407
Type 2 (Shadow/Different Material): 220
Type 3 (Shadow/Shadow, same material): 161
Type 4 (Everything Else): 229

The linear kernel and various types of nonlinear kernel functions (i.e. polynomial, sigmoid, RBF) were investigated. The RBF gave the best results, with the parameter search yielding optimal values of C=32 and γ=64. The accuracy was not very high: a five fold cross validation accuracy of 67.3%, with 370 support vectors. However, it must be kept in mind that the original accuracy was 92.55%, and now 67.3% of the remaining 7.45% misidentified boundaries could be accurately predicted.

The entire test set was revisited, but this time problem boundaries were first thresholded and treated separately. The original accuracy rate improved considerably[23]. In the next section, it is shown how shadow regions can be extracted from their corresponding boundaries.

[23] The accuracy improved from 92.55% to 96.4%. Theoretically a rate of 97.56% should have been achieved. After all, if 67.3% of the problem boundaries could be classified, and these boundaries accounted for 7.45% of the total, then there should have been an improvement of 5.01%. The reason that the accuracy rate was slightly lower was that 77.3%, not 100%, of the problem boundaries in the training set were thresholded. Thus, since it was not possible to threshold *all* of the problem boundaries, the accuracy were slightly (2%) lower than expected.

4.7 Extracting Shadow Regions From Their Boundaries

In this section a method for extracting shadow regions from their corresponding boundaries is described. Ideally, if $B_{ij} \cap R_i$ (boundaries B_{ij} that region R_i shares with its N neighbors $R_1 ... R_N$) are all identified as being shadow boundaries, R_i would be classified as being a shadow region. However, if a valid shadow boundary were to go undetected, then R_i would be incorrectly classified as being a non-shadow region. Therefore, in order to identify whether or not region R_i is truly a shadow region, the following two ratios are analyzed:

1. The proportion of the number of pixels in $B_{ij} \cap R_i$ that are in shadow to the total number of pixels in $B_{ij} \cap R_i$[24].
2. The proportion of the number of boundaries in $B_{ij} \cap R_i$ that are shadow boundaries to the total number of boundaries in $B_{ij} \cap R_i$[25].

If both of these ratios are sufficiently high then R_i is probably a shadow region. However, in order to avoid arbitrarily selecting thresholds for the two ratios, a third SVM was trained[26] to determine which regions were probable shadow regions. Finally, the test set used in Sections 4.5 and 4.6 was revisited, shadow boundaries were detected, and the corresponding shadow regions were extracted. (See Figure 4.14).

[24] For example, if a region shares 5 boundaries (of total size 100 pixels) with neighbors and two boundaries (of combined size 75) are in shadow, the ratio will be 75%.

[25] In this case, if a region shares 5 boundaries (of total size 100 pixels) with neighbors and two boundaries (of combined size 75) are in shadow, the ratio will be 40%.

[26] On a second test set of 50 images the SVM classified boundaries as being shadow or non-shadow, as described in Sections 4.5 and 4.6. Then, in each of the 50 images the actual shadow regions were manually identified. Each region for which the SVM detected at least one shadow boundary (1160 in our case) had the following written to a training file: the two ratios and whether or not it was actually a shadow region. The best results were found with the RBF, except this time the parameter search yielded optimal values of C=1 and γ=64. The total number of support vectors, 356 out of a possible 1160, indicated that over-fitting was not a problem, which was corroborated by a 5 fold cross-validation accuracy of 96.8%.

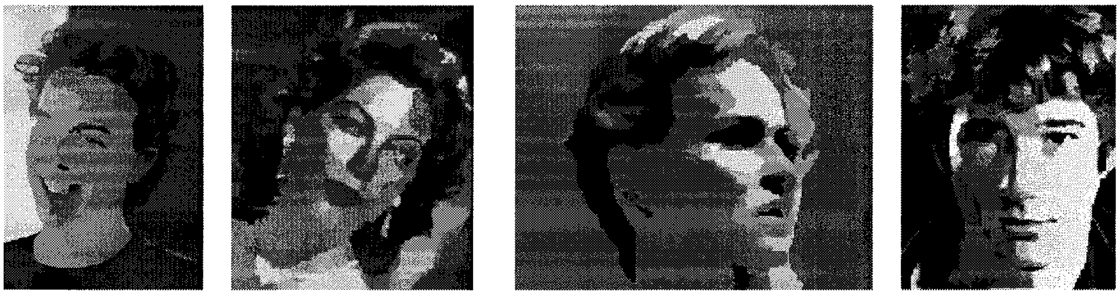


Figure 4.14 Extracting shadow regions from their boundaries: probable shadow regions are colored green. The original images, along with shadow boundary information, can be found in Figure 4.9.

The next step is to color the shadow regions, which is dealt with in Section 4.8.

4.8 Coloring Regions Inwards

Shadow boundaries were detected and categorized by the SVMs as per Table 1. This section describes how such boundary information can be used to color detected shadow regions with the average color of their non-shadow neighbors[27]. Firstly, adjacent shadow regions were merged into shadow super-regions[28]. Then, for a detected shadow super-region, boundaries with neighboring non-shadow regions were traversed, the average color along these boundaries was calculated[29], and the shadow super-region was assigned this color. Finally, all boundaries between shadow super-regions and neighboring non-shadow regions were smoothed[30] so that there was a uniform transition between shadow super-regions that had been colored in and neighboring non-shadow regions.

Results are shown in Figure 4.15. Overall, the method works well. However, as the shadows in the fourth image – and by extension their boundaries – are extremely strong, severe clipping at the lower end of the spectrum results in the loss or severe corruption of chromatic information. As discussed in Sections 4.5 and 4.6, the equations in Sections 4.2, 4.3 and 4.4 no longer hold in these cases. While treating these problematical boundaries separately dramatically improves the accuracy of their classification, they still suffer from a higher rate of misclassification than regular boundaries. As a result, the coloring process can be adversely affected, as seen in the fourth image of Figure 4.15.

[27] Henceforth, the term non-shadow neighbor refers to a non-shadow neighbor of the *same* material.
[28] A shadow often consists of several regions as discussed in Section 4.3 and illustrated in Figure 4.5. As the term shadow can be vague, the nomenclature shadow *super-region* will be used (where appropriate) to refer to a collection of adjacent shadow regions that belong to the same material.
[29] Color values along the boundary were taken from the output of the Luminance Retinex [6]
[30] Boundaries B_{ij} between shadow and non-shadow neighbors were smoothed by traversing them and convolving them with a Gaussian mask.

Figure 4.15 Coloring in shadows. **Top:** Shadows colored in with the average color of boundaries they share with non-shadow neighbors of the same material. **Bottom:** Smoothing applied to boundaries between shadow and non-shadow neighbors of the same material. The original images are shown in Figure 8.

4.9 Conclusions

The Introduction in Section 4.1 pointed out that most existing shadow detection and removal methods require prior information about the scene or impose unnatural constraints. Examples are knowledge of the illuminant vector [29] or that the scene be illuminated by white light [30-32]. The one algorithm in the literature that did not impose any unnatural constraints was the one proposed by Barnard and Finlayson [34]. However, as shown in this Sections 4.2-4.4 the theory was slightly flawed. As a result we proceeded to improve on this method and implemented the new technique

The shadow detection and removal technique proposed in this chapter does not require any camera calibration or other a priori information regarding the scene. It was found that Support Vector Machines were a powerful tool for identifying shadow boundaries based on their boundary properties. Furthermore, it was possible to use this boundary information to identify shadowed regions in the image and then assign them the color of non-shadow neighbors of the same material.

A primary goal of many statistical color image segmentation methods is to partition an image into regions, where each region corresponds to a particular material. Due to discontinuities resulting from both specularities and shadows, a given material in the scene may be segmented into several regions in the image. The next chapter shows how a more meaningful segmentation can be achieved after first compensating for illumination by detecting and removing specularities and shadows.

Chapter 5

Experiments and Results

5.1 Introduction

A primary goal of many color image segmentation methods [17,21-23,47,48] is to partition an image into regions, where each region corresponds to a particular material. Due to discontinuities resulting from shadows and specularities, a given material may be segmented into several regions, as illustrated by the facial skin in Figure 5.1.

Figure 5.1 Image segmentation. **Left:** Original, **Right:** Image segmented using EDISON.

In Section 5.2 it is shown that a more meaningful segmentation can be achieved after first compensating for illumination using the method proposed in this thesis. The method can be summarized as follows:

1. Apply the Luminance Retinex [6] in order to get dynamic range compression.
2. Detect and remove specularities as discussed in Chapter 3.
3. Detect and remove shadows as discussed in Chapter 4.

In Section 5.3 it is demonstrated that the accuracy of skin detection, a subset of color image segmentation, improves when this illumination compensation method is first applied. Finally, Section 5.4 shows how illumination compensation can increase the accuracy of face recognition.

5.2 Illumination Compensation: Results

The left-most image in Figure 5.1 was illumination compensated and then segmented. The results are shown in Figure 5.2.

Figure 5.2 Image segmentation after illumination compensation. **Left:** Left-most image from Figure 1 after illumination compensation, **Right:** Illumination compensated image after segmentation.

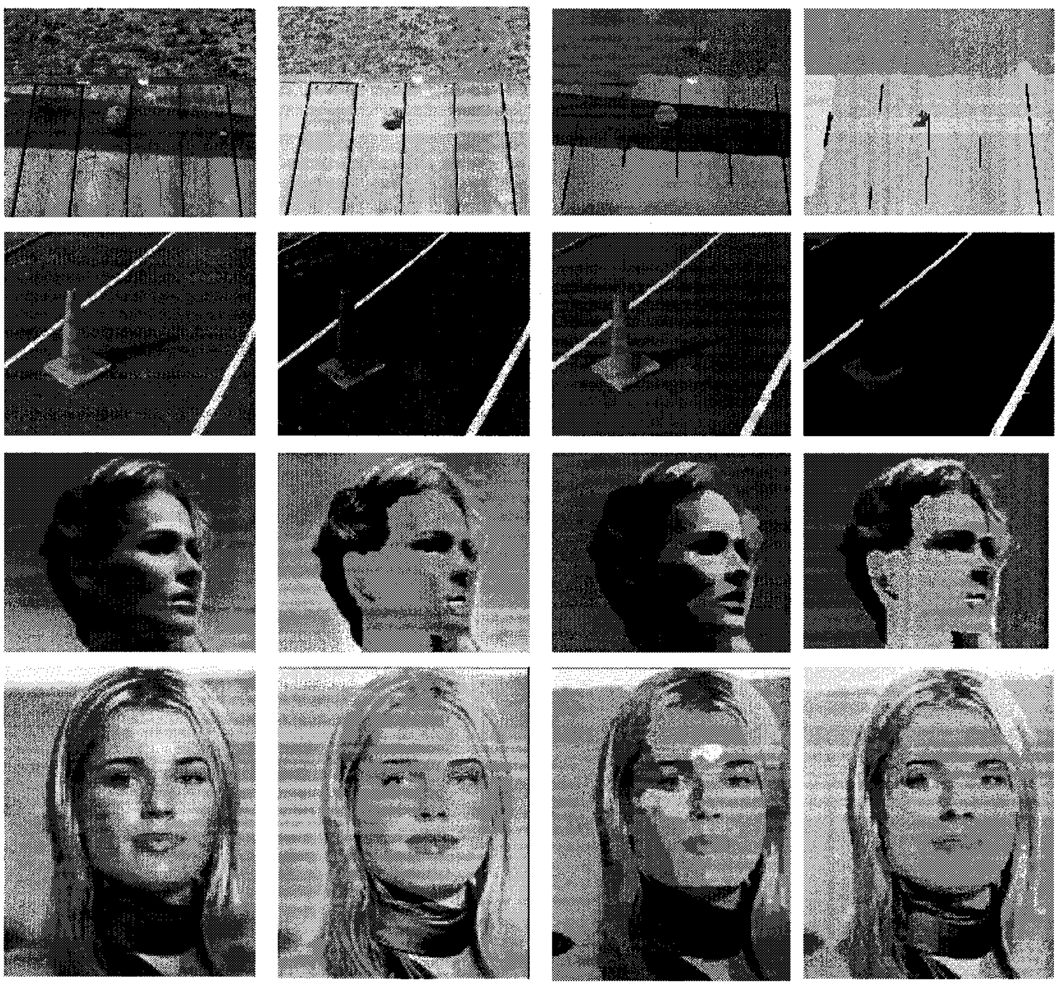
After shadows and specularities have been removed, the segmentation results become much more meaningful. Now the face will be segmented as a single region, as opposed to being divided into specularity, shadow, and non-shadow regions. Figure 5.3 shows results[31] of applying illumination compensation to a wide variety of images. The results show that after compensation, the image segmentations are not as adversely affected by discontinuities due to varying illumination. Figure 5.4 displays the same set of images after they have been processed by three other popular image enhancement techniques: the Multiscale Retinex with Color Restoration (MSRCR) [8], Histogram Equalization [84], and Gamma Correction [85] (with gamma = 2). All three have been applied in the literature to a wide variety of images .

In Figure 5.4 it can be seen that images enhanced by the MSRCR tend to be grayed out, as is also the case for images processed by Gamma Correction. Histogram Equalization, on the other hand, produces images that are fairly realistic. However, all three algorithms, fail to remove shadows and specularities as well as the method

[31] In Figure 5.3, the image in the second row is of a road that is in very bright sunlight. The compensated image is a bit darker due to the effect of the Luminance MSR, which lowered the exaggerated brightness of the original image by modifying the dynamic range. Thus, it is seen that not only does the Retinex brighten poorly lit images, it also lowers the brightness of overly lit images.

proposed in this thesis. In the next section it is demonstrated that compensating for illumination can improve skin detection in images.

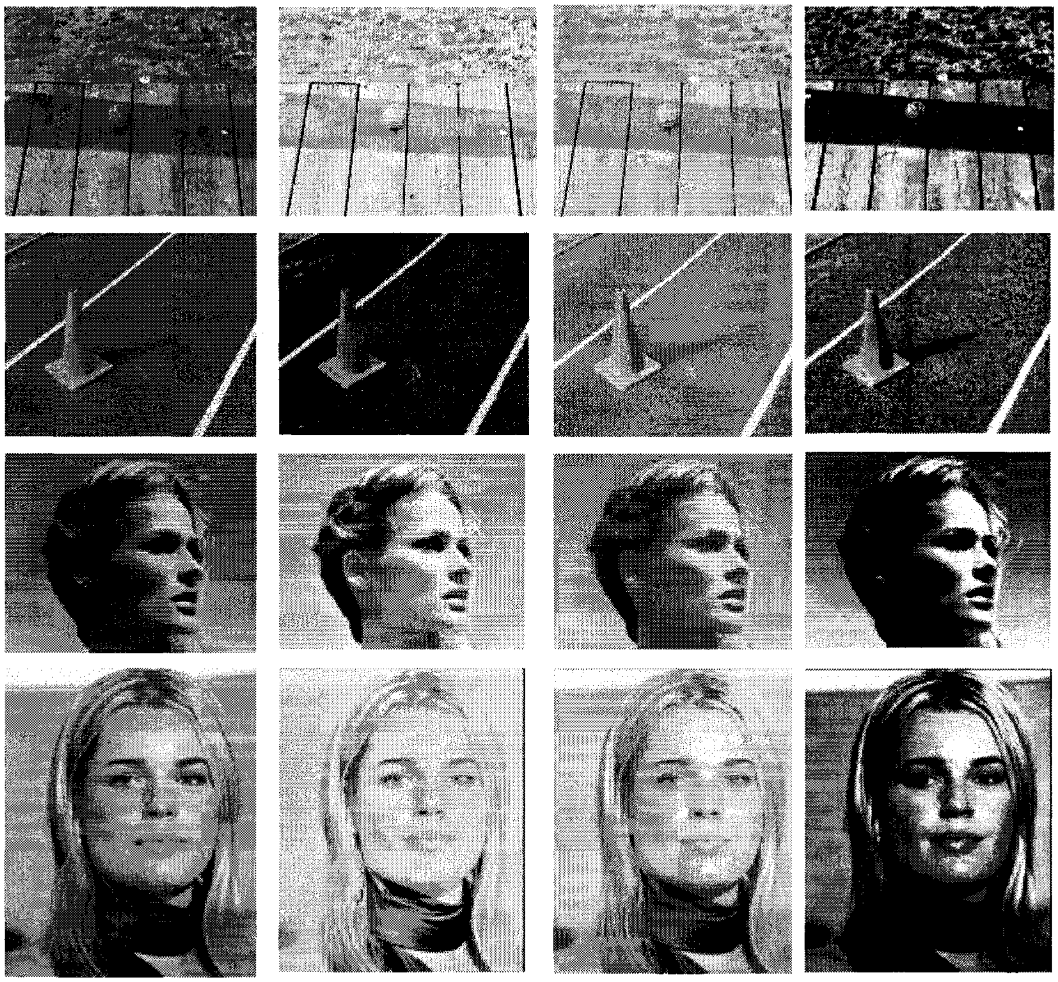
Original Image	Illumination Compensated Image	Segmented Original Image	Segmented Illumination Compensated Image



(a) (b) (c) (d)

Figure 5.3 Illumination compensation applied to a variety of images. From left to right: **(a)** Original image **(b)** Illumination compensated image **(c)** Original image segmented **(d)** Illumination compensated image segmented.

Original Image | **After MSRCR** | **After Gamma Correction** | **After Histogram Equalization**



(a) (b) (c) (d)

Figure 5.4 Various image enhancement techniques applied to the images in Figure 5.3. From left to right: **(a)** Original image **(b)** MSRCR **(c)** Gamma Correction **(d)** Histogram Equalization. Compare the images in this figure to the images in Figure 5.3 (b) which show the images processed by the algorithm in this thesis

5.3 Illumination Compensation for Skin Detection

Several image processing applications use skin detection to restrict the complexity of subsequent feature extraction. Applications range from face detection and tracking [62-65] to gesture recognition [66-68] and pornography filtering [48,69,70]. Skin detection techniques can be found in many commercial applications, for example the driver eye tracker developed by Ford UK [71].

Recent research has shown [35-41] that the skin color distribution under common indoor and outdoor illuminants falls in a shell-shaped region in chromaticity space that is close to the Planckian locus. This shell-shaped region, where skin can be found, is often referred to as the skin locus [35-41]. However, it is camera specific: Figure 5.4 shows the skin locus for two different cameras.

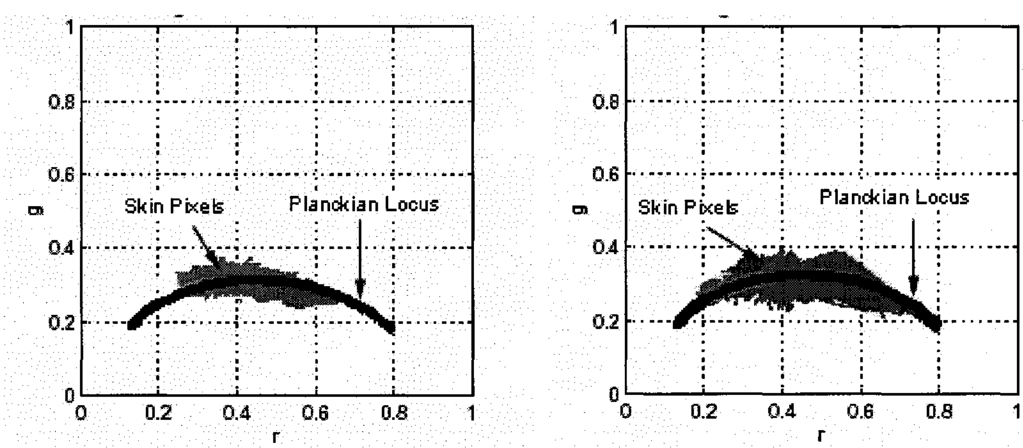


Figure 5.4 The Skin locus. Skin pixels are purple while the Planckian locus is the black curve. The skin locus is camera specific. **Left:** Nogatech camera [39], **Right:** Winnov camera [37]

Although the locus for each camera is slightly different, for the purposes of this section, the precise coordinates of the locus for any particular camera do not interest us. It is sufficient that for any given camera, skin pixels tend to cluster in a shell-shaped region around the Planckian locus. In this spirit, the loci from [35-41] are used as a basis for the creation of a *generic* skin locus[32], as illustrated in Figure 5.5. The generic skin locus can be used to create a simple skin detector: if a pixel is in the locus it is skin, and if it is outside the locus, it is not.

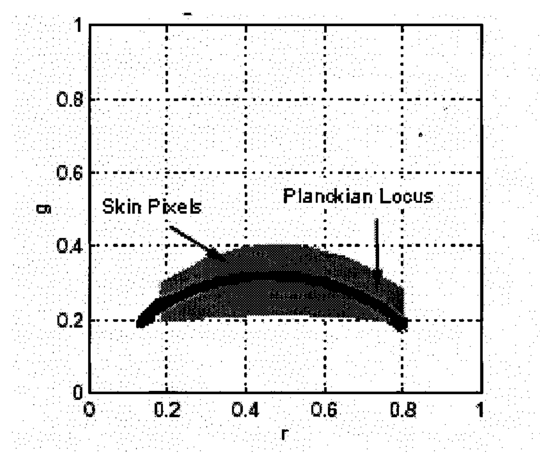


Figure 5.5 Generic Skin Locus. The locii of several cameras [17-23] were studied to create a skin locus that caters to a generic camera.

However, when skin is affected by strong shadows and specularities, skin pixels can fall outside the locus. Figures 5.6 and 5.7 demonstrate this more clearly. Figure 5.6 shows a facial image that suffers from strong lighting effects, alongside the illumination compensated version of the same image. Furthermore, Figure 5.6 also shows a manually obtained binary mask of the skin pixels of the face under consideration. Using this mask, skin pixels in both the original image and the illumination compensated image are plotted in Figure 5.7. Figure 5.7 shows that skin pixels in the original image do not always fall within the locus. Strong shadow and specularity cause many skin pixels to go astray and end up outside of the generic skin locus. After compensating for illumination, however, most of the wayward pixels return to the locus.

[32] The loci in [35-41] are merged to form the generic skin locus. Then, as in [39], a pair of quadratic functions are used to fit (in a least squares sense) the upper and lower bounds of the generic locus in rg chromaticity space. The upper bound quadratic function is $g=A_1r^2+B_1r+C_2$ and the lower bound quadratic function is $g=A_2r^2+B_2r+C_2$, where $A_1=-1.3571$, $B_1=1.3571$, $C_1=0.0893$, $A_2=-0.2857$, B_2,$=0.2857$, $C_2=0.1529$ and the horizontal axis is constrained to being between $r=0.2$ and $r=0.8$.

Figure 5.6 Binary mask of skin pixels. **Left:** Original, **Middle:** After illumination compensation, **Right:** Binary mask of skin pixels.

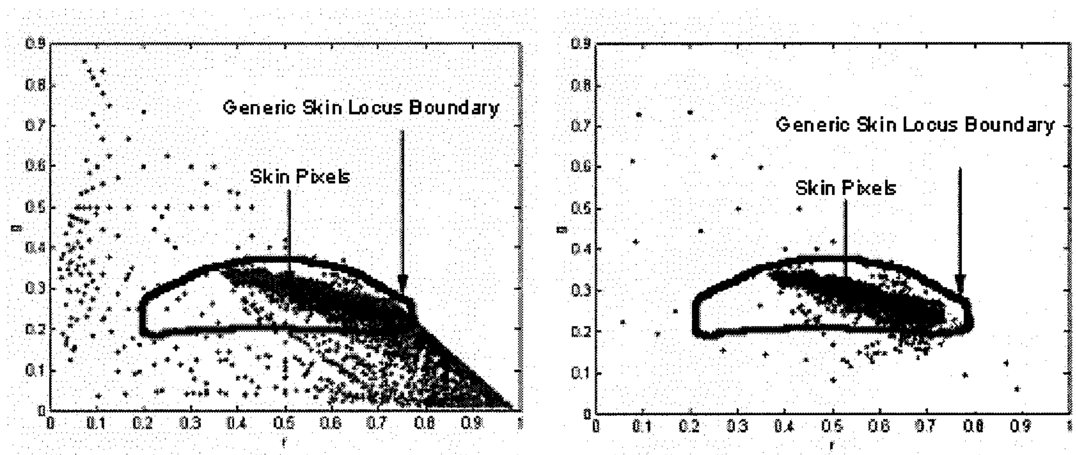


Figure 5.7 Distribution in chromaticity space of skin pixels within the binary mask. Skin pixels are in red and the boundary of the *generic* skin locus is black. **Left:** Original image, **Right:** After illumination compensation.

In order to better illustrate the results of Figure 5.7, in Figure 5.8 for both the uncompensated and compensated images we show an image of the face mask of Figure 5.6 with the labelled skin pixels in one color and the remaining errors in another.

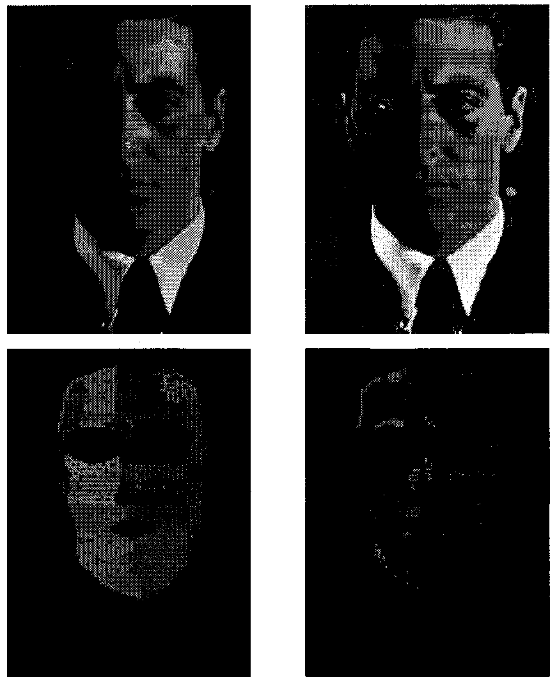


Figure 5.8 Face mask with correctly labeled skin pixels in red and the remaining errors in green. **Left:** Uncompensated image, **Right:** Illumination compensated image.

Figure 5.8 demonstrates that skin is far more accurately detected when illumination compensation is first applied to an image. If the skin locus for a specific camera is known, skin detection can be performed more accurately using that locus. Otherwise, when the camera is unknown, the generic skin locus described in this section can be used. In either case, by first removing specularities and shadows with illumination compensation, the likelihood that a skin pixel will in fact be detected as skin is increased.

5.4 Illumination Compensation for Face Recognition

Face recognition has a variety of applications, some of which include looking for missing children, law enforcement, and user authentication in order to restrict access to locations, equipment, and information. Recognition under varying illumination, however, can be challenging, as different lighting conditions often cause the same face to appear dramatically different [73-75]. This section describes experiments which show that the accuracy of face recognition can be improved if images are first compensated for illumination using the method proposed in this thesis. The face recognition experiments were performed on frontal images with varying illumination from the CMU PIE database[33] using a subspace analysis[34] technique called LNMF [76]. Figure 5.9 shows some of the images from the CMU PIE database that were used.

Figure 5.9 Typical frontal images with varying illumination from the CMU PIE database. Note the slight in-plane rotations, varying eye positions, and non-uniform background.

Not only do the faces in Figure 5.9 suffer from slight in-plane rotations, but they have varying eye positions and non-uniform backgrounds. All of these factors can adversely affect recognition rates. As the main focus of this thesis is illumination compensation, all

[33] The CMU PIE (Pose, Illumination, and Expression) database consists of 41,368 images of 68 people. Each person is imaged under 13 different poses, 43 different illumination conditions, and 4 different expressions. Experiments were only performed on images with frontal poses under varying illumination. There are 24 images of each person that fulfill the aforementioned criteria, for a total of 24x68, or 1632 images.

[34] Subspace methods [76-82] have become very popular in the field of face recognition. Typically, a set of training images from a face database are decomposed into a set of basis images. The images in the original training set are then represented as a linear combination of the N most siginificant basis images. An N dimensional feature space is then created from these bases, whereby each basis image is one of the dimensions. Finally, test images are projected into feature space, and if a test image occupies the same region of feature space as a training image, it is deemed likely that both images are of the same person.

face images were geometrically normalized[35] before they were presented to the LNMF recognition system. Figure 5.10 shows the results of normalizing the faces in Figure 5.9.

Figure 5.10 Images from Figure 5.9 after normalization.

After geometrical normalization, all images were converted to grayscale and then divided into training and test sets. As in [76], half of the images of each person were randomly assigned to a training set, while the test set consisted of the remaining images. Each image in the training set was then represented by LNMF as a linear combination of 81 basis images, and all training and test images were projected into the 81 dimensional feature space. For a given test image, the Euclidean distances from all training images were computed and the closest training image was deemed to be a match. The recognition accuracy[36] was found to be 93.5%[37]. The experimental process for this first experiment (referred to hereafter as Experiment 1) is illustrated in Figure 5.11.

[35] The normalized version of each face image satisfied the following constraints: (a) In-plane rotations were corrected by rotating images such that both eyes lay on a line that was parallel to the horizontal axis. (b) Images were resized such that the inter-ocular distance (distance between the eyes) was always 35 pixels. (c) Each image was cropped with an elliptical mask such that only the face was visible. All pixels outside the mask boundary were made black, thereby ensuring a uniform background for all images.

[36] The percentage of faces from the test set that are recognized correctly.

[37] In [83] PCA, LDA and FaceIt were applied to the same set of images, except in [83] gallery (known, or training) images were illuminated only by room lights, while probe (unknown, or test) images were illuminated by both room lights and the camera flash. The experiment in this thesis, however, randomly divided the images into training and test sets. In any case, the accuracy of PCA,LDA and FaceIt were 89%, 96% and 100%, respectively.

Experiment 1

All Training and Test Images→Geometrical Normalization

Input Test Image →Recognition

Experiment 2

All Training and Test Images→ Geometrical Normalization→Illumination Compensation

Input Test Image →Recognition

Figure 5.11 The experimental process: all images were normalized before recognition was performed.

The same experiment was repeated, but this time all training and test images were not only normalized, they were also compensated for illumination before recognition was performed (as depicted in Figure 10 under the heading Experiment 2: the accuracy increased to 98.4. Thus, illumination compensation improved the accuracy of face recognition by 4.9%.

5.5 Conclusions

It was found that a more meaningful segmentation could be achieved by compensating images for illumination using the method proposed in this thesis. Furthermore, the accuracy of skin detection, a subset of color image segmentation, was found to improve when this illumination compensation method was first applied. Finally, compensating images for illumination increased the accuracy of face recognition.

Chapter 6

Conclusions

This thesis proposes two new illumination invariant representations based on the Retinex: the $\hat{R}$ image, and the $\hat{R}_{EDGE}$ image. The $\hat{R}$ image is free of illumination effects, and the $\hat{R}_{EDGE}$ image is an edge image derived from the $\hat{R}$ image. Both representations can be used in conjunction with other computer vision applications. In this thesis the $\hat{R}_{EDGE}$ image was successfully used to detect material changes in an image as part of the proposed specularity detection process.

A novel approach to specularity detection and removal is also proposed in this thesis. The method does not require more than one camera or a sequence of images and no prior segmentation of the image is required. Furthermore, the technique is robust and can process complex scenes. The approach consists of finding the seed regions of specularities using saturation and intensity, upon which a wavefront is grown outwards either to the boundary of the specularity or until a material boundary is encountered in the newly discovered $\hat{R}_{EDGE}$ image. Once the boundary of the specularity has been reached, the wavefront moves inwards, coloring in as the region contracts. The process of coloring continues until the specularity no longer exists.

Also proposed in this thesis is a shadow detection and removal technique that does not require any camera calibration or other a priori information regarding the scene. The method uses Support Vector Machines to identify shadow boundaries based on their boundary properties. Shadowed regions are then inferred from these boundaries then assigned the color of non-shadow neighbors of the same material.

Illumination compensation as proposed in this thesis was found to increase the accuracy of image segmentation, skin detection and face recognition, all three of which are popular Computer Vision tasks with a wide range of applications.

However, the proposed method does not work with grayscale images. One advantage of storing database images as grayscale is that much less memory is required. While the method proposed in this thesis can process color images and those images can then be converted to grayscale, the method is powerless when given a grayscale image as input. As a result, popular databases such as the Yale face database cannot be processed with this method.

Another issue with the proposed technique is that the processed images often look artificial as a result of the coloring in process. Thus, the method should not be used for image enhancement purposes. Perhaps in the future this issue can be resolved by perfecting or introducing a completely new coloring in process.

While the method is relatively fast, it is not fast enough for real time. Each image takes approximately 3 to 4 minutes to process: about 30 seconds for the retinex, 1 minute for treating specularities, and 2 minutes for dealing with shadows. In the future it might be a good idea to implement the algorithm on a microchip, as this should result in much faster processing times.

This method did not remove all shadows in all images. It did, however, remove the most severe shadows most of the time. Similarly, the method did not remove all specularities in every image, but usually the strongest ones were treated by this method.

References

[1] G. D. Finlayson and S. D. Hordley, "Color Constancy at a Pixel," *Journal of the Optical Society of America A*, vol. 18, pp. 253-264, Feb. 2001.

[2] D. J. Jobson and G. A. Woodell, "Properties of a Center/Surround Retinex: Part 2. Surround Design," *NASA Technical Memorandum*, 1995.

[3] G. D. Finlayson and S. D. Hordley, "Color Constancy at a Pixel," *Journal of the Optical Society of America A*, vol. 18, pp. 253-264, Feb. 2001.

[4] P. B. Delahunt, "An evaluation of color constancy across illumination and mutual reflection changes," in *Department of Psychology*, UC Santa Barbara, Dec. 2001.

[5] G. Finlayson and G. Schaefer, "Single Surface Colour Constancy," *Proceedings of The 7th Color Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, USA, 1999.

[6] B. V. Funt, K. Barnard, M. Brockington, and V. Cardei, "Luminance-based multi-scale Retinex," *AIC'97, Kyoto, Japan*, May 1997.

[7] E. Land, "The Retinex Theory of Color Vision," *Scientific American*, pp. 108-129, Dec. 1977.

[8] E. Land, "Experiments in Color Vision," *Scientific American*, pp. 84-89, May 1959.

[9] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A Multi-Scale Retinex For Bridging the Gap Between Color Images and the Human Observation of Scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing*, vol. 6, pp. 965-976, July 1997.

[10] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multiscale Retinex for Color Rendition and Dynamic Range Compression," *Applications of Digital Image Processing XIX*, A. G. Tescher, ed., Proc. SPIE 2847, 1997.

[11] Z. Rahman, G. A. Woodell, and D. J. Jobson, "Retinex Image Enhancement: Application to Medical Images," *NASA Medical Imaging Conference*, July 2001.

[12] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Resiliency of the Multiscale Retinex Image Enhancement Algorithm," Proceedings of the IS&T Sixth Annual Color Conference, November 1998.

[13] Z. Rahman, 'Properties of a Center/Surround Retinex: Part 1. Signal Processing Design," *NASA Contractor Report 198194*, 1995 .

[14] D. J. Jobson and G. A. Woodell, "Properties of a Center/Surround Retinex: Part 2. Surround Design," *NASA Technical Memorandum*, 1995.

[15] Z. Rahman, G. A. Woodell, and D. J. Jobson, "A Comparison of the Multiscale Retinex With Other Image Enhancement Techniques," *Proceedings of the IS&T 50th Anniversary Conference*, May 1997.

[16] F. Torres, J. Angulo, and F. Ortiz, "Automatic Detection of Specular Reflectance in Colour Images Using the MS Diagram," *CAIP 2003*, pp. 132-139, 2003.

[17] G. J. Klinker, S. A. Shafer, and T. Kanade, "A phyisical approach to color image understanding,"" *International Journal of Computer Vision*, vol. 4, pp. 7-38, 1990.

[18] S. A. Shafer, "Using color to separate reflection components," *Color Research and Application*, vol. 10, pp. 210-218, 1985.

[19] K. Schlüns and A. Koschan, "Global and Local Highlight Analysis in Color Images," presented at Proc. 1st International Conference on Color in Graphics and Image Processing CGIP′2000, , Saint-Etienne, France, October 2000.

[20] G. J. Klinker, S. A. Shafer, and T. Kanade, "Image segmentation and reflection analysis through color," presented at Proc. IUW'88, Cambridge, Massachusetts, April 1988.

[21] R. Bajcsy, S. W. Lee, and A. Leonardis, "Image segmentation with detection of highlights and interreflections using color," Dept. of Computer and Info. Science, University of Pennsylvania, Technical Report 1989.

[22] R. Bajcsy, S. W. Lee, and A. Leonardis, "Color image segmentation and color constancy," *Proc. SPIE Perceiving, Measuring and Using Color*, vol. 1250, pp. pp. 245-255, 1990.

[23] R. Bajcsy, S. W. Lee, and A. Leonardis, "Color image segmentation with detection of highlights and local illumination induced by interreflections," presented at Proc. 10th Int. Conf. on Pattern Recognition 1990, Atlantic City, New Jersey, June 1990.

[24] S. Lin, Y. Li, and S. B. Kang, "Diffuse-Specular Separation and Depth Recovery from Image Sequences," presented at ECCV Lecture Notes in Computer Science, Springer-Verlag, 2002.

[25] A. Criminisi, S. B. Kang, and R. Swaminathan, "Extracting Layers and Analyzing Their Specular Properties Using Epipolar-Plane-Imaging Analysis," Columbia University 2002.

[26] K. Schluns and O. Wittig, "Photometric stereo for non-Lambertian surfaces using color information," presented at Proc. 7th Int. Conf. on Image Analysis and Processing, Capitolo, Monopoly, Italy, Sept. 20-22, 1993.

[27] S. W. Lee and R. Bajcsy, "Detection of specularity using color and multiple views," presented at Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy,, May 1992.

[28] C. Jiang and M. Ward, "Shadow Identification," *CVPR*, pp. 606-612, 1992.

[29] R. Charit and M. H. Loew, "Complex Shadow-Boundary Segmentation Using the Entry-Exit Method," *CVPR*, pp. 536-541, 1988.

[30] T. Gevers, "Adaptive Image Segmentation by Combining Photometric Invariant Region and Edge Information," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, 2002.

[31] T. Gevers and A. W. M. Smeulders, "Color Based Object Recognition," *Pattern Recognition*, vol. 32, pp. 453-464, March, 1999.

[32] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Shadow Identification and Classification Using Invariant Color Models," *ICASSP*, pp. 1545-1548, 2001.

[33] Graham D. Finlayson, Steven D. Hordley, Mark S. Drew, "Removing Shadows from Images," *ECCV* vol 4, pp. 823-836, 2002.

[34] K. Barnard and G. D. Finlayson, "Shadow Identification using Colour Ratios," *8th Color Imaging Conference*, pp. 97-101, 2000.

[35] M. Storring, H. Andersen, Skin colour detection under changing lighting condition. In *Araujo and J. Dias (ed.) 7th Symposium on Intelligent Robotics Systems*, 187–195, 1999.

[36] Martinkauppi, B., Soriano, M. ." Basis functions of the color signals of skin under different illuminants". In *3rd Intl conference on Multispectral Color Science*, pp. 21–24, 2001.

[37] Martinkauppi, B., Laaksonen, M., Soriano, M. "Behavior of skin color under varying illumination seen by different cameras at different color spaces". In *Machine Vision Applications in Industrial Inspection IX, Proceedings of SPIE*, vol. 4301, 102–112, 2001

[38] Soriano, M., Huovinen, S., Martinkauppi, B., Laaksonen, M. 2000. "Skin detection in video under changing illumination conditions". In *Proc. 15th International Conference on Pattern Recognition*, vol. 1, 839–842, 2000.

[39] M. Soriano, S. Huovinen, B. Martinkauppi, M. Laaksonen . "Using the skin locus to cope with changing illumination conditions in color-based face tracking". Proc. IEEE Nordic Signal Processing Symposium (NORSIG 2000), June 13-15, Kolmården, Sweden, pp. 383-386, 2000.

[40] M. Soriano, B. Martinkauppi, S. Huovinen, M. Laaksonen, "Skin color modeling under varying illumination conditions using the skin locus for selecting training pixels", Real-time Image Sequence Analysis Workshop (RISA2000), August 31-September 1, Oulu, Finland, 43-49 (2000).

[41] J. Fritsch, S. Lang, M. Kleinehagenbrock, G. A. Fink, G. Sagerer.,"Improving adaptive skin color segmentation by incorporating results from face detection", In Proc. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN), pages 337--343, Berlin, Germany, September 2002. IEEE.

[42] G. Ciocca, D. Marini, A. Rizzi, R. Schettini, and S. Zuffi, "On pre-filtering with Retinex in color image retrieval," presented at Proceedings of SPIE, 2001.

[43] G. Ciocca, D. Marini, A. Rizzi, R. Schettini, and S. Zuffi, "Color Based Image retrieval of uncalibrated images," presented at Proc. IEEE on Content-Based Multimedia Indexing, 2001.

[44] B. D. Thompson, Z. Rahman, and S. K. Park, "Retinex preprocessing for improved multi-spectral image classification," presented at Proc. SPIE 3716,, 1999.

[45] G. F. Miller and P. M. Todd, "A bottom-up approach with a clear view of the top: How human evolutionary psychology can inform adaptive behavior research. Review of J.H. Barkow, L. Cosmides, and J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture.," *Adaptive Behavior*, vol. 3, pp. 83-95, 1994.

[46] R. N. Shepard, "Evolution of a mesh between principles of the mind and regularities of the world.," in *The latest on the best: Essays on evolution and optimality*, J. Dupre, Ed. Cambridge, MA: MIT Press/Bradford Books, 1987.

[47] M. Celenk, "A Color Clustering Technique for Image Segmentation," *CVGIP*, pp. 145-170, 1990.

[48] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," Compaq 1998.

[49] P. W. Fieguth and S. Wesolkowski, "Highlight and Shading Invariant Color Image Segmentation Using Simulated Annealing," *EMMCVPR*, pp. 314-327, 2001.

[50] S. Wesolkowski, S. Tominaga, and R. D. Dony, "Shading and Highlight Invariant Color Image Segmentation," *Proc. of SPIE, Color Imaging: Device-Independent Color, Color Hard Copy, and Graphic Arts*, vol. 4300, pp. 229-240, 2001.

[51] A. Ögren, "Illumination and Shading Models," presented at Umeå University, Sweden.

[52] B. A. Maxwell and S. A. Shafer, "A Framework for Segmentation Using Physical Models of Image Formation," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR-94)*, June, 1994.

[53] Foley, V. Dam, Feiner, and Hughes, "Computer Graphics: Principles and Practice,," 2 ed: Addison Wesley, pp. p721-741, 760-766.

[54] C. B. Madsen, "Can Non-Existing Objects Cast Shadows?," presented at Proceedings: Eleventh Danish Conference on Pattern Recognition and Image Analysis, Copenhagen, Denmark, August 2002.

[55] J. S.-T. N. Cristianini, "An Introduction To Support Vector Machines," Cambridge University Press, 2000.

[56] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," 2003.

[57] C.-C. C. a. C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," presented at Department of Computer Science and Information Engineering, National Taiwan University, 2003.

[58] C.-C. C. C.-W. Hsu, C.-J. Lin, "The Analysis of Decomposition Methods for Support Vector Machines," presented at Department of Computer Science and Information Engineering, National Taiwan University, 1999.

[59] P. Meer and B. Georgescu, "Edge detection with embedded confidence," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23, pp. 1351-1365, 2001.

[60] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," presented at 16th International Conference on Pattern Recognition, Quebec City, Canada, August 2002.

[61] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, pp. 1351-1365, 2001.

[62] M. Soriano, S. Huovinen, B. Martinkauppi & M. Laaksonen (2000) Using the skin locus to cope with changing illumination conditions in color-based face tracking. Proc. IEEE Nordic Signal Processing Symposium (NORSIG2000), June 13-15, Kolmården, Sweden, 383-386 (2000).

[63] J. Fritsch, S. Lang, M. Kleinehagenbrock, G. A. Fink and G. Sagerer , Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection Proc. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN), pages 337--343, Berlin, Germany, September 2002. IEEE.

[64] R. L. Hsu, M. A. Mottaleb, and A. K. Jain, "Face detection in color images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696-706, May 2002.

[65] S. Kawato and J. Ohya, "Automatic skin-color distribution extraction for face detection and tracking," The 5th Intl. Conf. On Signal Processing, vol. II, pp. 1415-1418, Aug. 2000.

[66] MacLean WJ, Herpers R, Pantofaru C, Wood L, Derpanis K, Tsotsos JK. 'Fast Hand Gesture Recognition for Real-Time Teleconferencing Applications". *Second International Workshop on Recognition, Analysis and Tracking of Faces and*

Gestures in Real-time Systems, IEEE Computer Society, Vancouver, July 13-14 2001

[67] H. I. Christensen, D. Kragic and F. Sandberg, Computational Vision for Interaction with People and Robots, Centre for Autonomous Systems, Royal Institute of Technology, SE-100 44 Stockholm, Sweden

[68] Louise Clarke, Peter Harper, Richard B. Reilly, "Video Based Gesture Recognition for Augmentative Communication", *TIDE Conference*, Helsinki, Finland, July 1998

[69] Chan, Y. and Harvey, R.W. and Smith, D.J., "Building Systems to Block Pornography", *In BMVA conference*, 1999

[70] Margaret Fleck, David Forsyth, and Chris Bregler, "Finding Naked People," *Proceedings of the European Conference on Computer Vision,* B. Buxton and R. Cipolla, eds., Berlin, Germany: Springer-Verlag, Vol. 2, 1996, pp. 593-602

[71] D. Tock and I. Craw. Tracking and measuring drivers' eyes. *Image and Vision Computing*, 14:541–548, 1996.

[73] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for illumination changes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 19, pp. 721--732, July 1997.

[74] W. Zhao, R. Chellappa, A. Rosenfeld, and J. Phillips, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, pp. 399-458, 2003.

[75] A. S. Georghiades, D. J. Kriegman, and P. N. Belhumeur, "Illumination Cones for Recognition under Variable Lighting: Faces," presented at Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1998.

[76] T. Feng, S.Z. Li, H. Shum. H.J. Zhang. "Local Non-Negative Matrix Factorization as a Visual Representation". In *Proceedings of The 2nd International Conference on Development and Learning.* Washington DC. June, 2002.

[77] L. Sirovich and M. Kirby, ""Low-dimensional procedure for the characterization of human faces"," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, March 1987.

[78] M. Kirby and L. Sirovich, ""Application of the Karhunen- Loeve procedure for the characterization of human faces"," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, January 1990.

[79] Matthew A. Turk and Alex P. Pentland, ""Face recognition using eigenfaces."," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hawaii, June 1991, pp. 586-591.

[80] David Beymer, Amnon Shashua, and Tomaso Poggio, ""Example based image analysis and synthesis"," A. I. Memo 1431, MIT, 1993.

[81] A. P. Pentland, B. Moghaddam, and T. Starner, ""Viewbased and modular eigenspaces for face recognition"," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 84–91.

[82] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance", *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[83] Ralph Gross Jianbo Shi Jeff Cohn, "Quo vadis Face Recognition? The current state of the art in Face Recognition", Technical Report CMU-RI-TR-01-17 Robotics Institute, Carnegie Mellon University, Pittsburgh, PA

[84] John C. Russ, “The Image Processing Handbook”, 3rd Edition: CRC Press, October, 1998 pp. 233-242

[85] Foley, V. Dam, Feiner, and Hughes, "Computer Graphics: Principles and Practice," 2 ed: Addison Wesley, pp. 564-568