

Specularity Removal in Images and Videos: A PDE Approach

Satya P. Mallick¹, Todd Zickler², Peter N. Belhumeur³ and David J. Kriegman¹

¹ Computer Science and Engineering, University of California at San Diego, CA 92093

² Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

³ Computer Science, Columbia University, New York, NY 10027

Abstract. We present a unified framework for separating specular and diffuse reflection components in images and videos of textured scenes. This can be used for specularity removal and for independently processing, filtering, and recombining the two components. Beginning with a partial separation provided by an illumination-dependent color space, the challenge is to complete the separation using spatio-temporal information. This is accomplished by evolving a partial differential equation (PDE) that iteratively erodes the specular component at each pixel. A family of PDEs appropriate for differing image sources (still images vs. videos), differing prior information (e.g., highly vs. lightly textured scenes), or differing prior computations (e.g., optical flow) is introduced. In contrast to many other methods, explicit segmentation and/or manual intervention are not required. We present results on high-quality images and video acquired in the laboratory in addition to images taken from the Internet. Results on the latter demonstrate robustness to low dynamic range, JPEG artifacts, and lack of knowledge of illuminant color. Empirical comparison to physical removal of specularities using polarization is provided. Finally, an application termed *dichromatic editing* is presented in which the diffuse and the specular components are processed independently to produce a variety of visual effects.

1 Introduction

The reflectance of a wide variety of materials (including plastics, plant leaves, cloth, wood and human skin) can be described as a linear combination of specular and diffuse components. When this description is accurate, there are benefits to decomposing an image in this way. The diffuse reflectance component is often well-described by the Lambertian model, and by isolating this component, powerful Lambertian-based tools for tracking, reconstruction and recognition can be applied more successfully to real-world, non-Lambertian scenes. There is also evidence that specular reflectance plays a role in human perception, and there is a set of computer vision algorithms that rely solely on this component (e.g., [2, 5, 11]). Finally, in addition to image-analysis applications, specular/diffuse separation is important in image-based 3-D modeling, where (specular-free) diffuse texture maps are often desired, and in photo-editing, where the two components can be independently processed and recombined.

This paper addresses the separation of reflection components in images of general, possibly textured, scenes. We restrict our attention to surfaces that are well-represented

by Shafer’s dichromatic reflectance model [15], in which the spectral distribution of the specular component is similar to that of the illuminant while that of the diffuse component depends heavily on the material properties of the surface. The dichromatic model suggests the possibility of decomposing an image into its specular and diffuse components based on color information. Beginning with a single three-channel RGB image, the objective is to recover an RGB “diffuse image” and a monochromatic specular layer. This is an ill-posed problem, even when the illuminant color is known, and most existing methods operate by aggregating color information spatially across the image plane. We can differentiate between methods that are *global* and *local* in nature⁴.

Klinker et al. [6] show that when the diffuse color is the same at each point on an object’s surface, the color histogram of its image forms a T-shaped distribution, with the diffuse and specular pixels forming linear clusters. They use this information to estimate a single “global” diffuse color, and in principle, this approach can be extended to cases in which an image is segmented into several regions of homogeneous diffuse color. Results can be improved by exploiting knowledge of the illuminant color through transformations of color space [1, 18], but these methods also require an explicit segmentation of the scene into large regions of constant diffuse color. In recent work, R. Tan and Ikeuchi [16] avoid explicit segmentation by representing all of the diffuse colors in a scene by a global, low-dimensional, linear basis.

In addition to the global approaches mentioned above, there has been considerable interest in separating reflection components through purely local interactions. The advantage of this approach is that it admits highly textured scenes that do not contain piecewise constant diffuse colors. In most local methods, the illuminant color is assumed to be known *a priori*, which is not a severe restriction since it can often be estimated using established (global) methods (e.g., [7]). R. Tan and Ikeuchi [17] iteratively reduce the specular component of a pixel by considering one of its neighbors that putatively has a related diffuse component. P. Tan et al. [14] allow a user to specify a closed curve surrounding a specular region and then minimize an objective function based on local variations in diffuse chromaticity and specular intensity. One of the earliest local methods is that of Nayar et al. [10], which uses polarization as an additional cue to enable the recovery a spatially-varying source color.

The goal of this paper is to formalize the notion of “local interactions” for specular/diffuse separation, and thereby develop a general framework for achieving separation through local interactions in both images and videos. Unlike previous approaches⁵, the method is developed in the continuous domain, with local interactions governed by partial differential equations (PDEs). This process selectively shares color information between nearby image points through multi-scale erosion [3] with structuring sets that vary over the image plane. We derive a family of PDEs that are appropriate for differing conditions, including images of both textured and untextured surfaces. We also show

⁴ In addition to the color-based methods discussed here, there are a number of other methods that rely on multiple images and/or additional cues, such as variable lighting, variable polarization, and parametric reflectance. Readers are referred to [17] for a description of these methods.

⁵ A notable exception is the work of P. Tan et al. [14], who use a *variational* PDE to separate manually-segmented highlight regions. Our work differs in that it uses *morphological* PDEs enabling separation without the need for manual segmentation.

how this framework extends naturally to videos, where motion information is available as an additional cue.

On the practical front, this paper presents results on high-quality images acquired in the laboratory (Fig. 3a, 3b), and shows that they compare favorably to ground-truth determined using cross polarization (Fig. 4). Results on 8-bit images downloaded from the Internet (Fig. 3d, 3e) suggest robustness to artifacts caused by low dynamic range, JPEG compression, and lack of knowledge of the illuminant color. The paper also provides results on videos (Fig. 5) for which explicit optical flow is not necessarily available. Finally, an application – *dichromatic editing* – is presented (Fig. 6).

2 Background and Notation

The dichromatic model of reflectance is a common special case of the bidirectional reflectance distribution function (BRDF) model, and it was originally developed by Shafer [15] for dielectrics. According to this model, the BRDF can be decomposed into two additive components: the interface (specular) reflectance and the body (diffuse) reflectance. The model assumes that each component can be factored into a univariate function of wavelength and a multivariate function of imaging geometry, and that the index of refraction of the surface is constant over the visible spectrum. These assumptions lead to the following expression for the BRDF of a dichromatic surface:

$$f(\lambda, \Theta) = g_d(\lambda)f_d + f_s(\Theta), \quad (1)$$

where λ is the wavelength of light and $\Theta = (\theta_i, \phi_i, \theta_r, \phi_r)$ parameterizes the directions of incoming irradiance and outgoing radiance. The function g_d is referred to as the *spectral reflectance* and is an intrinsic property of the material. The functions f_d (constant for Lambertian surfaces) and f_s are the diffuse and specular BRDFs, respectively. Taking into account the spectral power distribution of a light source $L(\lambda)$ and a camera sensitivity function $C_k(\lambda)$, the image formation equation for a surface element with surface normal $\hat{\mathbf{n}}$, illuminated by a light source with direction $\hat{\mathbf{l}}$ is written

$$I_k = (D_k f_d + S_k f_s(\Theta)) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \quad (2)$$

$$\text{where } D_k = \int C_k(\lambda) L(\lambda) g_d(\lambda) d\lambda \text{ and } S_k = \int C_k(\lambda) L(\lambda) d\lambda.$$

An RGB color vector $\mathbf{I} = [I_1, I_2, I_3]^\top$ from a typical camera consists of three such measurements, each with a different sensitivity function with support in the visible spectrum. Note that S_k represents the effective source strength as measured by the k^{th} sensor channel and is independent of the surface being observed. Similarly, D_k is the effective albedo in the k^{th} channel. For notational simplicity, we define $\mathbf{S} = [S_1, S_2, S_3]^\top$ (with a corresponding definition for $\mathbf{D}$), and since scale can be absorbed by f_d and f_s , we assume $\|\mathbf{D}\| = \|\mathbf{S}\| = 1$.

3 Illuminant-Dependent Color Spaces

In the last few years there has been a burst of activity in defining color space transformations that exploit knowledge of the illuminant color to provide more direct access to the diffuse information in an image. While motivated by different applications, the transformations discussed here all share the same idea of linearly combining the three color channels of an RGB image to obtain one or two “diffuse channels”.

R. Tan and Ikeuchi [17] obtain a one-channel diffuse image through the transformation

$$I_d = \frac{3 \max_k (I_k/S_k) - \sum_k (I_k/S_k)}{3\tilde{\lambda} - 1}, \quad (3)$$

where $k \in \{1, 2, 3\}$, and the bounded quantity $1/3 < \tilde{\lambda} \leq 1$ is chosen arbitrarily. This transformation yields a positive monochromatic diffuse image, which can be seen by expanding Eq. 3 using Eq. 2 and assuming (for argument’s sake) that $I_1/S_1 > I_2/S_2, I_3/S_3$. In this case,

$$I_d = \frac{2I_1/S_1 - I_2/S_2 - I_3/S_3}{3\tilde{\lambda} - 1} = \frac{(2D_1/S_1 - D_2/S_2 - D_3/S_3) f_d \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}}{3\tilde{\lambda} - 1}. \quad (4)$$

Since this expression is independent of f_s and is directly related to $\hat{\mathbf{n}} \cdot \hat{\mathbf{l}}$, the positive image I_d is specular-free and depends directly on diffuse shading information.

An alternative transformation is proposed by Park [12], who isolates *two* predominantly diffuse channels while retaining a similarity to HSI color space. The transformation is composed of a linear transformation $\mathbf{L}_p$ and rotation $\mathbf{R}_p$, and is written

$$\mathbf{I}_p = \mathbf{R}_p \mathbf{L}_p \mathbf{I}, \quad \text{with} \quad \mathbf{R}_p \mathbf{L}_p \mathbf{S} = [0 \ 0 \ 2]^\top. \quad (5)$$

The matrices $\mathbf{R}_p$ and $\mathbf{L}_p$ are chosen such that the third color axis is aligned with the illumination color. As a result, that channel contains the majority of the specular component, leaving the other two channels to be predominantly diffuse.

A third transformation, proposed by Mallick et al. [8], defines a color space referred to as *SUV color space*. The transformation is written

$$\mathbf{I}_{SUV} = \mathbf{R} \mathbf{I}, \quad \text{with} \quad \mathbf{R} \mathbf{S} = [1 \ 0 \ 0]^\top. \quad (6)$$

Similar to Park’s transformation, one of the transformed axes in SUV space is aligned with the illuminant color. Unlike Park’s transformation, however, this channel includes *the complete* specular component, leaving the remaining two channels to be *purely diffuse*. To see this, we expand the expression for $\mathbf{I}_{SUV}$ using Eqs. 2 and 6 to obtain

$$\mathbf{I}_{SUV} = (\bar{\mathbf{D}} f_d + \bar{\mathbf{S}} f_s(\theta)) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \quad (7)$$

where $\bar{\mathbf{D}} = \mathbf{R} \mathbf{D}$ and $\bar{\mathbf{S}} = \mathbf{R} \mathbf{S} = [1, 0, 0]^\top$. Letting $\mathbf{r}_i^\top$ denote the i^{th} row of $\mathbf{R}$, the diffuse *UV* channels are

$$I_U = \mathbf{r}_2^\top \mathbf{D} f_d \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \quad I_V = \mathbf{r}_3^\top \mathbf{D} f_d \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \quad (8)$$

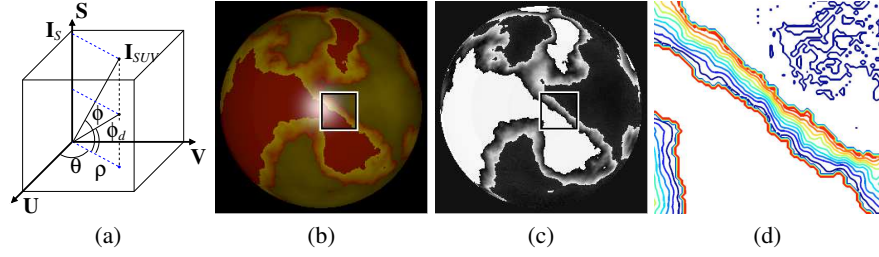


Fig. 1. (a) A color in the SUV color space is parameterized by (ρ, θ, ϕ) . ρ and θ are independent of specularity, and θ generalizes the notion of “hue” for arbitrarily colored illuminants. The problem of removing specularity is reduced to finding ϕ_d , the diffuse part of ϕ . (b) A rendered RGB image of a textured sphere. (c) The value of θ at each pixel of the image. Notice that θ is constant in regions of constant diffuse color and is independent of specularity as well as shading. (d) Blown-up view of the iso-contours of θ in the rectangular region indicated in (b) and (c). White indicates regions of constant θ . In textured images, erosion of the specular component occurs along iso-contours of θ , which ensures that diffuse texture is preserved while the specularity is removed.

which depend only on diffuse-shading and are specular-free. The S -channel is given by

$$I_S = \mathbf{r}_1^\top \mathbf{D} f_d \hat{\mathbf{n}} \cdot \hat{\mathbf{l}} + f_s(\theta) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}. \quad (9)$$

It contains all of the specular component in addition to an unknown diffuse component.

Each of the three transformations described in this section exploits knowledge of the illuminant to provide a partial dichromatic separation, which is an important step toward our stated goal. Of the three, the SUV color space defined in Eq. 6 is the best-suited for our purpose. Unlike Eq. 3, it is a linear transformation that yields two “diffuse” channels, and unlike Eq. 5, these two “diffuse” channels are in fact completely free of specularity. As described in the next section, these properties lead to a generalized notion of hue that can be used as a guide for local interactions, enabling the computation of a complete specular/diffuse separation even in cases of significant diffuse texture.

4 Specularity Removal and Differential Morphology

This section derives a family of non-linear PDEs for completing the partial specular/diffuse separation provided by a transformation to SUV color space. Intuitively, these PDEs define a series of local interactions in which color information is shared along curves (or surfaces) of constant “hue.”

We begin by re-parameterizing SUV color space using a combination of cylindrical and spherical coordinates. As depicted in Fig. 1, suppressing the spatial dependence for notational simplicity, we define

$$\rho = \sqrt{I_U^2 + I_V^2}, \quad \theta = \tan^{-1} \left(\frac{I_U}{I_V} \right), \quad \phi = \tan^{-1} \left(\frac{I_S}{\rho} \right). \quad (10)$$

This parameterization has the following properties:

1. Since they depend only on the diffuse UV channels, both ρ and θ are independent of the specular reflectance component.
2. Since the illuminant color is aligned with the S -axis, the angle θ parameterizes the pencil of dichromatic planes in an image. We refer to θ as *generalized hue*, since it reduces to the standard definition of hue in the special case of a white illuminant. It depends on the direction of the diffuse color vector but not the magnitude of the diffuse component.
3. ρ represents diffuse shading, since it is directly related to $\hat{\mathbf{n}} \cdot \hat{\mathbf{l}}$, and therefore, the magnitude of the diffuse component.
4. ϕ is a linear combination of specular and diffuse components, and we can write $\phi = \phi_s + \phi_d$, where ϕ_s and ϕ_d are the specular and diffuse contributions to ϕ .

According to these properties, the problem of computing a specular/diffuse separation is reduced to one of estimating $\phi_d(x, y)$, the diffuse contribution to ϕ at each image point. Once the scalar function $\phi_d(x, y)$ is known, the RGB diffuse component follow directly from inverting the transformations in Eqs. 10 and 6, with ϕ replaced by ϕ_d .

4.1 Multi-scale Erosion

Our goal is to compute a specular/diffuse separation through estimation of the scalar function $\phi_d(x, y)$ through purely local interactions. This section describes how this can be accomplished by evolving a PDE that iteratively “erodes” the specular contribution to ϕ and converges to an estimate of ϕ_d at each point. The erosion process is guided locally by the diffuse color information provided by ρ and θ , and is formulated in the continuous domain using one of a family of non-linear PDEs that define multi-scale erosion [3]. The theory presented in this section is related to the formulation of multi-scale erosion presented by Brockett and Maragos [3].

The multi-scale erosion $\varepsilon(\mathbf{x}, t)$ of a bivariate function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ by structuring set $B \subseteq \mathbb{R}^2$ at scale t is defined as

$$\varepsilon(\mathbf{x}, t) = (f \ominus tB)(\mathbf{x}) \triangleq \inf\{f(\mathbf{x} + \Delta\mathbf{x}) : \Delta\mathbf{x} \in tB\},$$

where the set B is compact, and $tB \triangleq \{tb : b \in B\}$. Intuitively, $\varepsilon(\mathbf{x}, t)$ evaluated at a particular value of t corresponds to an erosion of the function $f(\mathbf{x})$, where the function value at $\mathbf{x} = (x, y)$ is replaced by the minimum of all function values in the “neighborhood” tB , which is a scaled replica of structuring set B . A multi-scale erosion is computed by considering the PDE

$$\frac{\partial \varepsilon}{\partial t}(\mathbf{x}, t) = \lim_{\Delta t \rightarrow 0} \frac{\varepsilon(\mathbf{x}, t + \Delta t) - \varepsilon(\mathbf{x}, t)}{\Delta t}. \quad (11)$$

When the structuring set is both compact and convex, the multi-scale erosion has a semigroup structure, allowing one to write [3]

$$\frac{\partial \varepsilon}{\partial t}(\mathbf{x}, t) = \lim_{\Delta t \rightarrow 0} \frac{\inf\{\nabla \varepsilon^\top \Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t B\}}{\Delta t}, \quad (12)$$

	Textureless	Texture 1	Texture 2	Video 1	Video 2
Set	2D Disk	2D Line	2D Ellipse	3D Disk	3D Line
Direction	Isotropic	Iso-cont. of θ	Iso-cont. of θ	Iso-surf. of θ	Optic Flow
$\mathbf{M}$	$\mathbf{I}_{2 \times 2}$	$\mathbf{I}_{2 \times 2} - \nabla \hat{\theta} \nabla \hat{\theta}^\top$	$\mathbf{A} \mathbf{A}^\top$	$\mathbf{I}_{3 \times 3} - \nabla \hat{\theta} \nabla \hat{\theta}^\top$	$\mathbf{F} \mathbf{F}^\top / \ \mathbf{F}\ ^2$

Fig. 2. Summary of five cases from left to right: (1) image with uniform diffuse color, (2-3) textured image, (4) video, and (5) video with known optical flow. Rows depict: the structuring set used, the direction/surface of erosion, and the matrix $\mathbf{M}$ in the multi-scale erosion equation (Eq. 14.) $\mathbf{I}_{n \times n}$ is the identity matrix, and $\nabla \hat{\theta}$, $\mathbf{A}$ and $\mathbf{F}$ are as defined in Sec. 4.

where $\nabla \varepsilon$ is the two-dimensional spatial gradient of ε evaluated at t . Finally, as shown in [3], in the special case where B is disk-shaped, Eq. 12 becomes

$$\varepsilon_t = -\|\nabla \varepsilon\|. \quad (13)$$

Eq. 13 is an example of a PDE that can be used for specular/diffuse separation, albeit in the special case when the scene consists of a texture-less surface with uniform diffuse color. To see this, suppose we are given an input image with corresponding functions $\rho(\mathbf{x})$, $\theta(\mathbf{x})$ and $\phi(\mathbf{x})$, and suppose we define $\varepsilon(\mathbf{x}, 0) = \phi(\mathbf{x})$. The solution to Eq. 13 evaluated at scale t corresponds to the erosion of ϕ by a disk-shaped structuring set, meaning that the value of ϕ at each image point is replaced by the minimum value within a disk-shaped neighborhood of radius t . Since $\phi_d(\mathbf{x}) \leq \phi(\mathbf{x})$, it follows that when the image contains at least one image point that is purely diffuse (that is, for which $\phi_s = 0$) then $\varepsilon(\mathbf{x}, t)$ evaluated at t will converge to $\phi_d(\mathbf{x})$ as t is made sufficiently large. In the next three sub-sections, we develop more sophisticated PDEs for cases of multiple regions of uniform diffuse color, complex diffuse texture, and video. In all of these, the basic idea is the same: the value of ϕ_d at each image point is estimated by eroding the initial function ϕ . By changing the structuring set, however, the process can be controlled so that region boundaries and diffuse texture are preserved during the process. In particular, we show that the PDE governing the evolution of ϕ for three different cases – texture-less images, textured images, and video – can all be written as

$$\varepsilon_t = -g(\rho, \nabla \rho) (\nabla \varepsilon^\top \mathbf{M} \nabla \varepsilon)^{1/2}, \quad (14)$$

where $\mathbf{M}$ is a different matrix for each case. $g(\rho, \nabla \rho)$ is called the stopping function and is defined in the following section. Fig. 2 summarizes the cases we consider.

4.2 Texture-less Surfaces: Isotropic Erosion

Eq. 13 describes a process in which the specular component of ϕ is eroded equally in all directions. This is desirable in cases of homogeneous diffuse color, but if regions of distinct color exist, there is a possibility that “color bleeding” may occur. To prevent this, we introduce a “stopping function” analogous to that used in anisotropic diffusion [13]. A stopping function is useful for attenuating the erosion process in two different cases

1. If a region of the surface is “white” (i.e., it reflects all wavelengths equally) or if the surface is the same color as the light source, the diffuse component of color cannot be isolated. Since $\rho = 0$ in this case, no diffuse color information is available, and erosion should be arrested.
2. Information about ϕ should not be shared across boundaries between regions of distinct color. Since these boundaries often coincide with large values of $\|\nabla\rho\|$, erosion should be attenuated when $\|\nabla\rho\|$ is large.

One possible stopping function that meets these guidelines is

$$g(\rho, \nabla\rho) = \left(\frac{1 - e^{-\rho}}{1 + e^{-\rho}} \right) \frac{e^{-(\|\nabla\rho\| - \tau)}}{1 + e^{-(\|\nabla\rho\| - \tau)}}, \quad (15)$$

where τ is a threshold on $\|\nabla\rho\|$, above which erosion is heavily attenuated. Incorporating this into Eq. 13 yields

$$\varepsilon_t = -g(\rho, \nabla\rho) \|\nabla\varepsilon\| = -g(\rho, \nabla\rho) (\nabla\varepsilon^\top \mathbf{I}_{2 \times 2} \nabla\varepsilon)^{1/2}. \quad (16)$$

The erosion process defined by this equation can be used for the specular/diffuse separation of images containing large regions of uniform diffuse color.

4.3 Textured Surfaces: Anisotropic Erosion

An example of a scene that does not contain regions of uniform diffuse color is shown in Fig. 1 (b). In this case, eroding the function ϕ isotropically would blur the diffuse texture. Instead, we need to erode ϕ anisotropically, only sharing information between neighboring image points for which ϕ_d is likely to be equal. Of course, we have no information about the diffuse color *a priori*, so it is impossible to know the correct neighborhood (if it even exists) with certainty. As depicted in Fig. 1 (c, d), since θ is independent of both specular and shading information, the directions tangent to the iso-contours of $\theta(\mathbf{x})$ provide a good choice. In the absence of any additional information, they provide a good local predictor for the direction in which ϕ_d is constant.

We define

$$\nabla\hat{\theta} = \begin{cases} \nabla\theta/\|\nabla\theta\| & \|\nabla\theta\| > 0 \\ 0 & \|\nabla\theta\| = 0, \end{cases} \quad (17)$$

where $\nabla(\cdot)$ refers to the spatial gradient, and we denote the direction orthogonal to $\nabla\theta$ by $\mathbf{V}$.⁶ The multi-scale erosion of ϕ with the spatially-varying, linear structuring sets $\mathbf{V}(\mathbf{x})$ is derived analogous to the isotropic (disk-shaped) case discussed previously.

$$\begin{aligned} \varepsilon_t &= \lim_{\Delta t \rightarrow 0} \frac{\inf\{\nabla\varepsilon^\top \Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t \mathbf{V}\}}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{-\Delta t |\nabla\varepsilon^\top \mathbf{V}|}{\Delta t} = -|\nabla\varepsilon^\top \mathbf{V}|. \end{aligned} \quad (18)$$

⁶ Since θ is periodic, a definition of distance is necessary for its gradient to be correctly computed. We define the distance between two angles θ_1 and θ_2 as $\min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|)$.

Using the fact that $\mathbf{V} = [\hat{\theta}_y \ -\hat{\theta}_x]^\top$ (or $[\hat{\theta}_y \ -\hat{\theta}_x]^\top$), and including the stopping function, we obtain

$$\varepsilon_t = -g(\rho, \nabla\rho) \left[\nabla\varepsilon^\top \left(\mathbf{I}_{2 \times 2} - \nabla\hat{\theta}\nabla\hat{\theta}^\top \right) \nabla\varepsilon \right]^{1/2}. \quad (19)$$

Using similar arguments to that in the isotropic case, it can be shown that $\varepsilon(\mathbf{x}, t)$ evaluated at sufficiently large t will be equal to $\phi_d(\mathbf{x})$ (and will yield a correct specular/diffuse separation) if the iso-contour of θ passing through each point $\mathbf{x}$: 1) contains only points for which ϕ_d is constant; and 2) contains at least one point at which a purely diffuse observation ($\phi_s = 0$) is available. Note that in regions where the diffuse color is constant (i.e., $\nabla\hat{\theta} = [0 \ 0]^\top$), this equation reduces to Eq. 16, and the erosion becomes isotropic as desired.

In practice, the transition from linear to disk-shaped structuring sets in Eq. 19 is controlled by a threshold on $\|\nabla\theta\|$. This discontinuous transition can be avoided by employing an elliptical structuring set with a minor axis aligned with the direction of $\nabla\theta$ and with an eccentricity that varies smoothly with $\|\nabla\theta\|$. To derive a PDE for the corresponding multi-scale erosion, we let E denote an elliptical structuring set, and we describe this set by the lengths of its major and minor axes (λ_1, λ_2) and the angle between its major axis and the x -axis (ψ). Points $\mathbf{x}$ on the boundary of E satisfy $\mathbf{x}^\top \mathbf{Q} \mathbf{x} = 1$ where $\mathbf{Q} = \mathcal{R}(-\psi) \mathbf{\Lambda}^{-2} \mathcal{R}(\psi)$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2)$ and $\mathcal{R}(\psi)$ is a clockwise rotation of the plane. As before, the multi-scale erosion defined by this set satisfies

$$\varepsilon_t = \lim_{\Delta t \rightarrow 0} \frac{\inf\{\nabla\varepsilon^\top \Delta \mathbf{x} : \Delta \mathbf{x} \in \Delta t E\}}{\Delta t}. \quad (20)$$

To simplify the right-hand side of this equation, we define the transformation $\mathbf{x} = \mathbf{A} \mathbf{x}'$, with $\mathbf{A} = \mathcal{R}(-\psi) \mathbf{\Lambda} \mathcal{R}(\psi)'$. The spatial gradient of ε with respect to $\mathbf{x}'$ is then given by the chain rule: $\nabla\varepsilon' = \mathbf{A}^\top \nabla\varepsilon$. The transformation $\mathbf{A}$ maps the set E to the unit disk (since $\mathbf{x}^\top \mathbf{Q} \mathbf{x} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{Q} \mathbf{A} \mathbf{x} = \mathbf{x}'^\top \mathbf{x}' = 1$), and as a result, we can write $\inf\{\nabla\varepsilon^\top \Delta \mathbf{x} : \Delta \mathbf{x} \in \Delta t E\} = \inf\{\nabla\varepsilon'^\top \Delta \mathbf{x}' : \Delta \mathbf{x}' \in \Delta t B\}$. Substituting this into Eq. 20 and comparing with Eq. 13, we obtain $\varepsilon_t = -\|\nabla\varepsilon'\| = -\left(\nabla\varepsilon^\top \mathbf{A} \mathbf{A}^\top \nabla\varepsilon\right)^{1/2}$. Finally, the addition of the stopping function yields

$$\varepsilon_t = -g(\rho, \nabla\rho) \left(\nabla\varepsilon^\top \mathbf{A} \mathbf{A}^\top \nabla\varepsilon \right)^{1/2}. \quad (21)$$

4.4 Videos: Anisotropic Erosion in Three Dimensions

Thus far, we have dealt exclusively with still images, but the framework extends naturally to video, which can be treated as a 3D volume $\mathbf{I}(x, y, z)$ in which time is the third dimension (z). As in the case of textured images, the direction of $\nabla\theta$ is assumed to be a good local predictor for the direction (in 3D space-time) of maximum diffuse color change. We would like to preserve the component of $\nabla\phi$ along this direction during the erosion process, which is accomplished by restricting the erosion of ϕ to the iso-surfaces of θ . In the absence of additional information, there is no preferred direction within an iso-surface of θ , so a natural choice of structuring set is a circular disk contained within its tangent plane.

To compute the multi-scale erosion equation, we note that the structuring set described above consists of a disk (denoted C) whose surface normal is aligned with $\nabla\theta$. Thus, the maximum projection of $\nabla\phi$ onto the plane that contains this disk is given by $\left(\|\nabla\phi\|^2 - \|\nabla\hat{\theta}^\top \nabla\phi\|^2\right)^{1/2}$, and the evolution equation can be simply written as

$$\begin{aligned}\varepsilon_t &= \lim_{\Delta t \rightarrow 0} \frac{\inf\{\nabla\varepsilon^\top \Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t C\}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{-\Delta t \left(\|\nabla\varepsilon\|^2 - \|\nabla\hat{\theta}^\top \nabla\varepsilon\|^2\right)^{1/2}}{\Delta t} \\ &= -\left(\|\nabla\varepsilon\|^2 - \|\nabla\hat{\theta}^\top \nabla\varepsilon\|^2\right)^{1/2}, \text{ where } \nabla\varepsilon = [\varepsilon_x \ \varepsilon_y \ \varepsilon_z]^\top.\end{aligned}$$

After some algebraic manipulations, and incorporating the stopping function, we obtain

$$\varepsilon_t = -g(\rho, \nabla\rho) \left[\nabla\varepsilon^\top (\mathbf{I}_{3 \times 3} - \nabla\hat{\theta}\nabla\hat{\theta}^\top) \nabla\varepsilon \right]^{1/2}. \quad (22)$$

Note that the erosion equation for textured and texture-less surfaces are special cases of the erosion equation for videos.

As mentioned earlier, if some *a priori* information is known, better structuring sets can be designed. An interesting example is when optical flow estimates are available at each location in a video. We let $[u(x, y, z) \ v(x, y, z)]^\top$ represent the estimated optical flow at location (x, y, z) in the video, so that space-time points (x, y, z) and $(x + u, y + v, z + 1)$ correspond to projections of the same surface element. It follows that ϕ_d can be estimated by eroding ϕ along the direction $\mathbf{F} = [u \ v \ 1]^\top$. Using the expression for erosion by a linear set derived in Eq. 19 we obtain

$$\varepsilon_t = -g(\rho, \nabla\rho) \left| \frac{\mathbf{F}^\top}{\|\mathbf{F}\|} \nabla\varepsilon \right| = -g(\rho, \nabla\rho) \left(\nabla\varepsilon^\top \frac{\mathbf{F}\mathbf{F}^\top}{\|\mathbf{F}\|^2} \nabla\varepsilon \right)^{1/2}. \quad (23)$$

5 Results

The methods were evaluated using images and videos acquired in the laboratory as well as those downloaded from the Internet. Using a known (or approximately known) illuminant color, each image is transformed into SUV space, and the functions ρ , θ and ϕ are computed. Specular/diffuse separation is achieved by numerically evolving the appropriate multi-scale erosion PDE with initial condition $\varepsilon(\mathbf{x}, 0) = \phi(\mathbf{x})$. The process is complete when the maximum change in ε is below a selected threshold, and this yields an estimate of $\phi_d(\mathbf{x})$, which completely defines the specular and diffuse components.

It is important to note that the non-linear PDEs governing erosion are defined at points where the partial derivatives exist. Even if this is satisfied by the initial data, however, at finite scales a multi-scale erosion generally develops discontinuities referred to as *shocks*. Shocks can be dealt with (as we do here) by replacing standard derivatives by morphological derivatives [3]. They can also be handled using viscosity solutions [4].

Fig. 3 (a, b) shows two 12-bit images⁷ acquired in a controlled setting (with known illuminant color) along with the recovered specular and diffuse components. Both re-

⁷ All images in this section should be viewed on a monitor or high-quality color print. The images can be viewed at a higher resolution by zooming into the PDF document.

sults were obtained using the anisotropic erosion defined in Eq. 19. The method correctly handles both regions of uniform color (e.g., the orange pepper Fig. 3 (a)) and regions with significant texture (e.g., the pear in Fig. 3 (b)). Looking closely at the pear, we notice that diffuse texture that is barely visible in the input image is revealed when the specularity is removed. Figure 3 (c) shows a 12-bit image of a human face in which the illuminant color was unknown and was assumed to be white. Again, diffuse texture is preserved, while the specular component is reduced. Pixels on the forehead between the eyebrows are saturated, and therefore violate the dichromatic model. The stopping function (Eq. 15) ensures that these pixels are implicitly identified and treated as outliers during the erosion process. While left here for illustrative purposes, these artifacts can be reduced by inpainting the diffuse and/or specular components in a post-process. (This is done, for example, by P. Tan et al. [14].)

Figure 4 compares the result of our algorithm with the ground truth obtained using polarization filters on the light source and camera. The polarizer in front of the light source is fixed while the polarizer in front of the camera is rotated to an orientation that produces an image with minimum specularity. The result of our algorithm is very close to the ground truth on both the textured surfaces (i.e., the vase and pear) and the untextured surfaces (i.e., the sphere).

Additional still-image results are shown in Fig. 3 (d, e). These images were downloaded from the Internet, so they exhibit low dynamic range (8-bit) and are corrupted by JPEG compression. Since illuminant color was not known, it was assumed to be white, and the gamma was assumed to be 2.2. Despite these sources of noise, the multi-scale erosion defined in Eq. 19 still succeeds in separating the diffuse and specular components. An animation of the erosion process accompanies this paper.

In addition to still images, we also evaluated the method on video sequences, some frames of which are shown in Fig. 5. In both cases, erosion is performed along iso-surfaces of θ using Eq. 22, and in both cases, texture is preserved while the specularity is removed. Complete videos accompany this paper.

5.1 Dichromatic Editing

To further demonstrate the efficacy of our approach, we use it as a means for *dichromatic editing* – the simulation of visual effects by the independent processing of reflection components. Some examples are shown in Fig. 6, where: 1) the specular and diffuse components are recovered using Eq. 19, 2) each component is processed individually, and 3) they are recombined. Since the diffuse and specular components often form two distinct components of visual perception, dichromatic editing can achieve a variety of visual effects, including the effects of make-up, surface roughening, and wetness.

6 Conclusion

This paper presents a framework for specular/diffuse separation in images and video that is based on local spatial (and spatio-temporal) interactions. Separation is framed in terms of differential morphology, which leads to a family of non-linear PDEs. By evolving these PDEs, we effectively erode the specular component at each image point.

This erosion is guided by local color and shading information, so that diffuse texture is preserved without requiring an explicit segmentation of the image. By developing the problem in terms of morphological PDEs, we can benefit from existing robust numerical algorithms to solve them [9], which is an important advantage over purely discrete formulations. In addition, videos are naturally considered in this formulation, with the erosion equation for videos including the still-image equations as a special case.

The approach described in this paper relies purely on local color information, and is therefore limited to dichromatic surfaces for which the diffuse and specular colors are distinct. It requires the illuminant color to be known (at least approximately) *a priori*. In the future, we plan to overcome these limitations by exploiting additional cues, such as local shape, in addition to color.

7 Acknowledgment

The authors would like to thank Roger Brockett for helpful comments regarding the theoretical development in Sec. 4. This work was supported in part by the National Science Foundation. S. Mallick and D. Kriegman were funded under IIS-03-08185 and EIA-02-24431, T. Zickler was funded under IIS-05-41173, and P. Belhumeur was funded under IIS-03-08185, EIA-02-24431, and ITR-00-85864.

References

1. R. Bajcsy, S. Lee, and A. Leonardis. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *IJCV*, 17(3):241–272, March 1996.
2. A. Blake and G. Brelstaff. Geometry from specularities. In *ICCV*, pages 394–403, 1988.
3. R. Brockett and P. Maragos. Evolution equations for continuous scale morphology. *IEEE trans. on Sig. Proc.*, 42:3377–3386, 1994.
4. M. G. Crandall, H. Ishii, and P. L. Lions. *User’s Guide to Viscosity Solutions of Second Order Partial Differential Equations*, volume 26. July 1992.
5. G. Healey and T. O. Binford. Local shape from specularity. *CVGIP*, 42(1):62–86, 1988.
6. G. Klinker, S. Shafer, and T. Kanade. The measurement of highlights in color images. *IJCV*, 2(1):7–32, 1988.
7. H. S. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *JOSAA*, 3(10):1694–1699, 1986.
8. S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond Lambert: Reconstructing specular surfaces using color. In *CVPR*, pages II: 619–626, 2005.
9. P. Maragos and M. A. Butt. Curve evolution and differential morphology. *Fundamenta Informaticae*, 41:91–129, 2000.
10. S. Nayar, X. Fang, and T. Boulton. Separation of reflection components using color and polarization. *Int. Journal of Computer Vision*, 21(3):163–186, 1997.
11. M. Osadchy, D. Jacobs, and R. Ramamoorthi. Using specularities for recognition. In *ICCV*, pages 1512–1519, 2003.
12. J. B. Park. Efficient color representation for image segmentation under nonwhite illumination. In *SPIE, Volume 5267, pp. 163-174 (2003)*.
13. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *PAMI*, 12(7):629–639, 1990.

14. T. Ping, S. Lin, L. Quan, and H.-Y. Shum. Highlight removal by illumination-constrained inpainting. In *ICCV*, pages 164–169, Nice, France, 2003.
15. S. Shafer. Using color to separate reflection components. *COLOR research and applications*, 10(4):210–218, 1985.
16. R. Tan and K. Ikeuchi. Reflection components decomposition of textured surfaces using linear basis functions. In *CVPR*, pages I: 125–131, 2005.
17. R. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *PAMI*, 27(2):178–193, February 2005.
18. T. Tan, R. K. Nishino, and K. Ikeuchi. Separating reflection components based on chromaticity and noise analysis. *PAMI*, 26(10):1373–1381, October 2004.

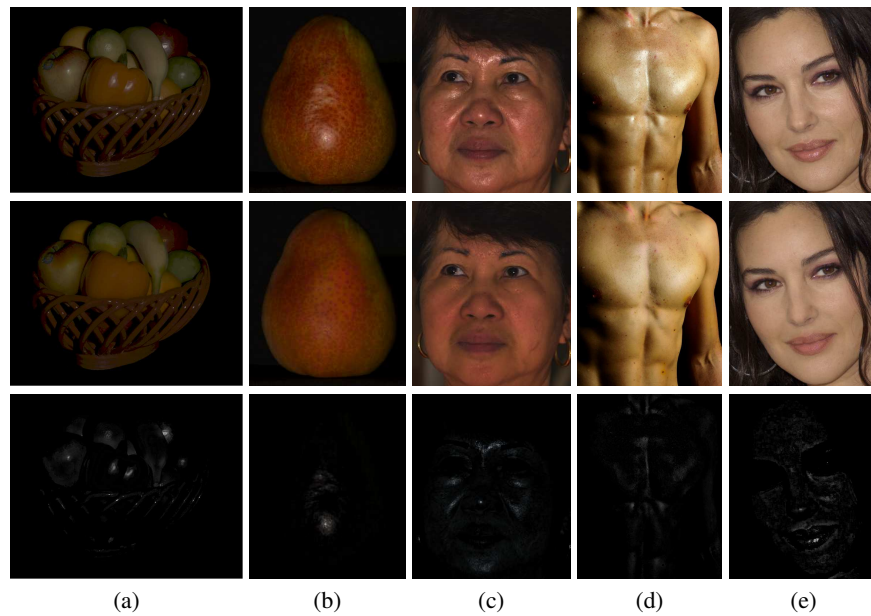


Fig. 3. Separation results for images. Top row: Input images. Middle row: Diffuse Component. Bottom row: Specular component. Equation 19 is used in all cases, since it naturally handles both textured and untextured surfaces. The 12-bit input images in (a, b) were acquired in the laboratory under known illuminant color. In (c), the illuminant color was not known and was assumed to be white. 8-bit JPEG images (d, e) were downloaded from the Internet, the illuminant was assumed to be white, and the gamma was assumed to be 2.2. Despite these sources of noise, diffuse and specular components are successfully recovered.



Fig. 4. Comparison to ground truth. Left: input image. Center: ground truth diffuse component obtained using linear polarizers. Right: diffuse component recovered using anisotropic multi-scale erosion.

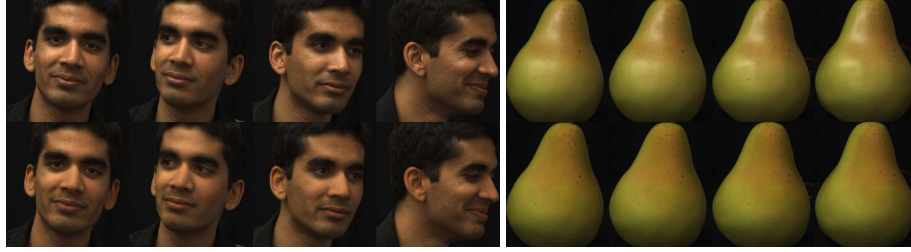


Fig. 5. Separation results for (12-bit) video. Top row: Frames from input sequences. Bottom row: Diffuse component recovered using Eq. 22. (Complete videos accompany this paper.)

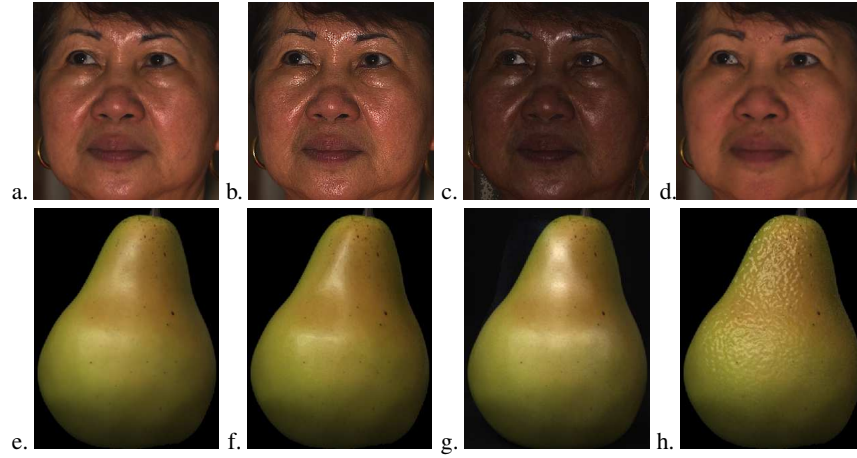


Fig. 6. Dichromatic editing examples. In each case a visual effect is simulated by independent processing of the recovered specular and diffuse components. (a) Input image. (b) Wetness effect by sharpening the specular component. (c) Skin color change by varying the intensity of the diffuse component. (d) Effect of make-up by smoothing the diffuse component and removing the specular component. (e) Input image. (f) Sharpened specular lobe, as would occur if the surface was more smooth. This is achieved by eroding the specular component using a disk-shaped structuring element and amplifying it. (g) Effect of an additional light source obtained by exploiting the object symmetry and reflecting the specular component about the vertical axis. (h) Avocado-like appearance by modulating the specular component.