Analysis and synthesis of multicolored objects in a single image

Hui-Liang Shen and John H. Xin

Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong, China

Received December 8, 2004; accepted May 20, 2005

We present a method with which to recover the intrinsic shading and reflectance characteristics of multicolored three-dimensional objects in a single image, with which realistic new scenes can be synthesized. A color watershed algorithm, which is based on a regularized dichromatic fitting error, is proposed for robust image segmentation. For shading recovery in small regions, a weighted interpolation is employed, whereas in large regions the reflectance and shading are calculated based on the assumption of gradual shape variation. It is demonstrated that the proposed method is promising and can be applied in image simulation.

OCIS codes: 100.3020, 330.1710, 110.2960.

The recovery and reconstruction of a three-dimensional object shape from an image are among the most challenging problems in optical imaging and computer vision.[1] In the literature, techniques have been proposed to recover surface properties by use of a set of images or a single image. For example, some methods have been proposed to recover both shading and reflectance properties from a single image by assuming that an abrupt change in chromaticity is more likely caused by reflectance than by shape.[2,3] Funt *et al.*[2] detected reflectance changes in chromaticity channels and recovered shading in the thresholded gradient of a luminance channel by using a Fourier transform. As the color information alone may not be sufficient, Tappen *et al.*[3] employed a classifier to distinguish image derivatives caused by material boundary or shape variation. A limitation of these two methods is that they assume perfect diffuse surfaces. Tian and Tsui[4] introduced a technique to recover the shape of a non-Lambertian uniformly colored surface from a single image exhibiting both diffuse and specular reflections. They developed a reflectance map of a hybrid surface under an extended light source and recovered the surface's shape by using a shape-from-shading technique. However, their approach cannot be applied when the objects are multicolored.

In this Letter we propose a method first to recover intrinsic shading and reflectance characteristics of multicolored three-dimensional objects from a single image and then to use these recovered characteristics to synthesize new scenes. The calculation of the proposed method is carried out in red–green–blue (RGB) space, and thus the reflectance mentioned in this Letter is also represented by RGB values. In comparison with the technique described in Refs. 2 and 3, our method considers both diffuse and specular reflections and requires the accurate decomposition of shading and reflectance for the sake of new scene synthesis. Similarly to that in Ref. 4, our method is also based on the dichromatic reflection model,[5] considering its accuracy in describing the interaction between light and surface for the majority of inhomogeneous materials.[6,7] The proposed method consists of three parts. First, a new color watershed algorithm is proposed for image segmentation based on the regularized dichromatic fitting error. Then the intrinsic shading of the object is recovered by the consideration of both small and large component regions. Finally, based on the recovered geometry, a new object scene can be accurately synthesized.

For optical inhomogeneous materials, the interaction between light source and object surface can be described by a dichromatic reflection model.[5] In the model, the total light reflection is a combination of body and surface reflections. These two reflections can each be further decomposed into color and geometry components. More precisely, the color $\mathbf{V}^p$ at pixel position p is the linear combination of intrinsic body color $\mathbf{K}$ and surface color $\mathbf{H}$ in RGB space:

$$\mathbf{V}^p = \alpha^p \mathbf{K} + \beta^p \mathbf{H} = \mathbf{M}\mathbf{c}^p, \tag{1}$$

where α^p and β^p are the geometric coefficients of diffuse and surface reflections, respectively, $\mathbf{V}^p = [V_1^p\ V_2^p\ V_3^p]^{\mathrm{T}}$, $\mathbf{K} = [K_1\ K_2\ K_3]^{\mathrm{T}}$, $\mathbf{H}^p = [H_1\ H_2\ H_3]^{\mathrm{T}}$, $\mathbf{M} = [\mathbf{K}\ \mathbf{H}]$, and $\mathbf{c}^p = [\alpha^p\ \beta^p]^{\mathrm{T}}$, where T is the vector transpose. It is assumed in this Letter that $\mathbf{H}$ is the same as the illumination color and is known *a priori*. Without loss of generality, it can be assumed that the illumination color is white. If $H_1 \neq H_2 \neq H_3$, the color of each pixel p can easily be normalized to be CV_i^p/H_i, where C is a scale constant.

To recover shading and reflectance, one should separate the multicolored object into individual regions with uniform color. Klinker *et al.* proposed a segmentation method for understanding an image scene by combining the dichromatic reflection model and the characteristics of cameras.[8] In this Letter, as the accuracy of the segmentation result is of more concern, we suggest using a watershed algorithm. In the well-known algorithm proposed by Vincent and Soille,[9] the gray-scale image is regarded as a topographic relief consisting of catchment basins and watershed lines. The final segmented regions are formed by immersion simulation. The main advantage of a watershed algorithm is that it includes not only region integrality but accuracy in boundary localization as well. In this Letter we extend the watershed algorithm to three-dimensional color space, which is referred to as a color watershed hereafter. To overcome the inherent problem of oversegmentation of the watershed algorithm, we select seeds for each region. For this purpose we calculate the hue

$$h^p = \arctan\left[\frac{\sqrt{3}(V_2^p - V_1^p)}{2V_3^p - V_1^p - V_2^p}\right] = \arctan\left[\frac{\sqrt{3}(K_2^p - K_1^p)}{2K_3^p - K_1^p - K_2^p}\right]. \tag{2}$$

In Eq. (2), hue is geometry invariant and is related only to the intrinsic body color of each region. This indicates that one can decide regions and their hue ranges from a hue histogram by detecting significant peaks. For the ith region, let the hue with the maximum number of pixels be the region's representative hue, $\bar{h}_i$. Then pixels whose hue values satisfy $|h^p - \bar{h}_i| \leq \delta$, where δ is a threshold, are selected as the seeds of the ith region. For each region, the median color of the seed pixels is approximately regarded as body color $\mathbf{K}_i$ in the segmentation process. To obtain accurate segmentation results in boundaries, we consider both global and local information in the color watershed algorithm. It uses the regularized fitting error of a dichromatic model as the distance measurement for each pixel with respect to each region:

$$d_i^p = \|\mathbf{V}^p - \mathbf{M}_i(\mathbf{M}_i^+\mathbf{V}^p)\| + \lambda|[\mathbf{M}_i^+\mathbf{V}^p]_1 - \bar{\alpha}|, \tag{3}$$

where $\mathbf{M}_i = [\mathbf{K}_i\ \mathbf{H}]$, $\mathbf{M}_i^+$ is the pseudoinverse of $\mathbf{M}_i$, $\bar{\alpha}$ is the mean α value of the already labeled neighboring pixels of p, λ is a weight-controlling shape variation, and $[\mathbf{M}_i^+\mathbf{V}^p]_1$ represents the first term of vector $\mathbf{M}_i^+\mathbf{V}^p$, which is actually the α of the ith region. In Eq. (3), the former term expresses the global characteristic of a pixel by using the direct fitting error with the least-squares criterion. For some pixels, this fitting error may be small for more than one region. Therefore it is necessary to take into account the local information, i.e., the slow shape variation, which is expressed by the last term in Eq. (3). In the color watershed algorithm we make use of a data structure termed a sequentially sorted list. The pixel is inserted into the sequentially sorted list in ascending order with respect to the minimum fitting error $d^p = \min_i d_i^p$. In each flood process or iteration, a pixel p with minimum d^p is popped out for labeling. If the labels of the neighboring pixels that were already labeled in previous iterations are the same, pixel p is simply assigned this label. Otherwise, the pixel is assigned a label with the minimum d_i^p according to Eq. (3).

After color watershed segmentation, the reflectance image is represented by the body colors, and the shading image is represented by geometric coefficient α. Note that α values of the large and main regions may be accurate, whereas those of small, narrow, or dark regions may not be so reliable because of imaging noise and blurring effects in regional boundaries. Therefore, for pixels in these unreliable regions, one should interpolate their colors by using those of large neighboring regions. In this Letter, one-dimensional polynomial interpolations conducted on four directions (horizontal, vertical, and two diagonal) are described, and the final interpolated color is their weighted sum.

For two large neighboring regions, we consider geometric coefficient $\mathbf{c}^p$ reliable. However, as body color $\mathbf{K}_i$ is represented by the median color of the seed pixels, the recovered geometry will not be continuous in boundaries. Therefore, to obtain true shading, one should reestimate the body colors of regions. In this Letter we assume that the normal directions of neighboring pixels are the same and hence that so are their $\mathbf{c}^p$ values. This assumption is reasonable, as the shape variation is always gradual. We let the largest region be the reference region, with median color $\mathbf{K}_0$ as its body color. We select a neighboring region, say region n, with the longest watershed. Then we search the pixel pairs beside the watershed in these two regions and let the geometric coefficient of the pixel pair be equal. More specifically, we let α_0^p and β_0^p be the geometric coefficients of the ith pixel pair of the reference region and $\mathbf{V}_{n,i}^p$ be the corresponding color of region n. Then, for each pixel pair, we have the following equation:

$$\alpha_0^p\mathbf{K}_n + \beta_0^p\mathbf{H} = \mathbf{V}_{n,i}^p. \tag{4}$$

Body color $\mathbf{K}_n$ can easily be estimated by use of all the pixel pairs along the watershed based on the least-squares method, and the geometric coefficients $\mathbf{c}^p$ of the pixels inside region n can be recalculated. Using this technique, we can sequentially estimate the accurate body color and geometry for the remaining regions. With the recovered shading and reflectance, new three-dimensional objects can be easily synthesized by use of Eq. (1).

According to our previous studies,[6,7] the color difference between an actual image and a synthesized image is small in both spectral reflectance space and RGB space. Therefore in this Letter we do not repeat the quantitative analysis but evaluate the performance of the method based on image observation. Figure 1 shows the image segmentation results of us-

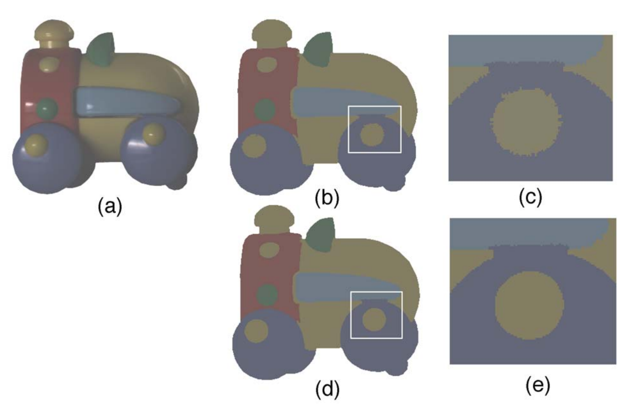

Fig. 1. (Color online) Results of segmentation of the color watershed algorithm on (a) the image of a toy train with parameters (b) $\lambda = 0$ and (d) $\lambda = 1$. (c), (e) Details of the rectangles in (b) and (d), respectively.

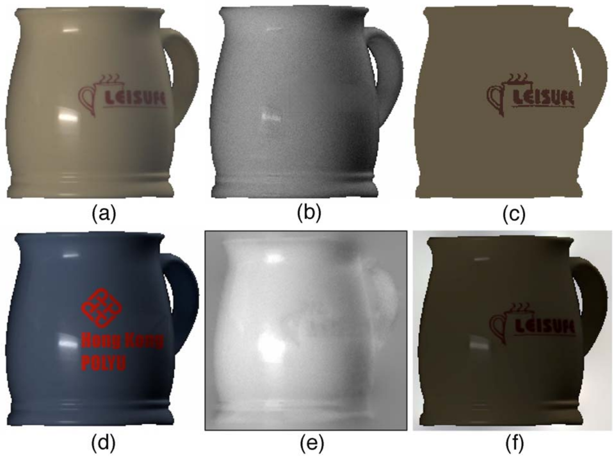

Fig. 2. (Color online) Experimental results for the image of a cup: (a) original image; (b), (e) shading images recovered with the proposed method and with Tappen's method, respectively; (c), (f) reflectance images recovered with the proposed method and with Tappen's method, respectively; (d) new image synthesized by the proposed method.

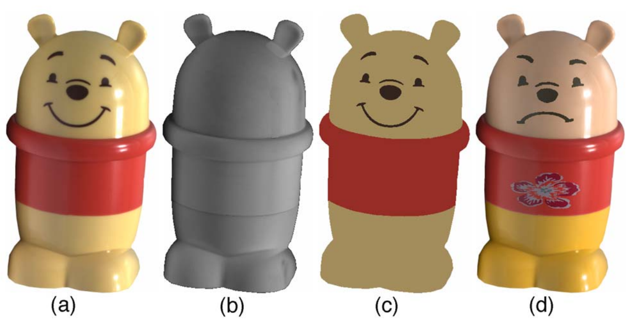

Fig. 3. (Color online) Experimental results for a toy image: (a) original image, (b) shading image, (c) reflectance image, (d) synthesized new image.

ing the proposed color watershed algorithm with different λ values. As expected, the segmentation is more accurate when one is considering the constraint of slow shape variation, especially in the low-intensity areas. In the investigation we found that $\lambda=1$ is suitable for most images. Figure 2 illustrates an example of the separation of shading and reflectance images, as well as showing the new synthesized images. Using the proposed method, we reliably recovered the shading in the small writing regions, using the weighted polynomial interpolation. As the shading or α image is calculated according to Eq. (1) by the standard least-squares method, it does not seems smooth. However, this appearance will not affect the following synthesis process, as discussed in our previous studies.[6,7] This conclusion is proved by the natural and photorealistic appearance of the synthesized red pattern in Fig. 2(d). On the contrary, it is seen that the results of Tappen's method[3] are not so good. It is obvious that the shading [Fig. 2(e)] in the highlighted regions is not reliably recovered and that the reflectance [Fig. 2(f)] contains a clear shape variation at the bottom right-hand side. Figure 3 shows another example. The image in Fig. 3 contains both small and large regions, and they were separately treated by the techniques discussed above. The specular reflection of the object is strong and is beyond the dynamic range of the camera. These small highlighted regions are also compensated for by the polynomial interpolation. It can be seen from the synthesized images that the technique developed is promising.

In conclusion, our aim is to recover the intrinsic shading and reflectance characteristics of a multicolored object in a single image and to synthesize new scenes afterward. First the image is separated by use of a color watershed algorithm based on a regularized fitting error of the dichromatic reflection model. Then, for small regions, as the geometry may not be accurate, we obtain the color by using a weighted polynomial interpolation in four directions. For large regions, the body color is calculated based on the assumption that the normal directions of neighboring pixel pairs are the same. After recovery of the shading image, new three-dimensional objects can be synthesized. We have demonstrated that the proposed method is promising and can be applied in image simulation.

The authors acknowledge financial support from the Hong Kong Polytechnic University. H.-L. Shen's e-mail address is tcshenhl@polyu.edu.hk.

References

1. Y. Sando, M. Itoh, and T. Yatagai, Opt. Lett. **28**, 2518 (2003).
2. B. V. Funt, M. S. Drew, and M. Brockington, in *ECCV-92: Second European Conference on Computer Vision*, G. Sandini, ed. (Springer-Verlag, 1992), pp. 124–132.
3. M. F. Tappen, W. T. Freeman, and E. H. Adelson, in *Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference*, S. Becker, S. Thrun, and K. Obermayer, eds. (MIT Press, 2003), p. 1343.
4. Y. L. Tian and H. T. Tsui, J. Opt. Soc. Am. A **14**, 397 (1997).
5. S. A. Shafer, Color Res. Appl. **10**, 210 (1985).
6. J. H. Xin and H. L. Shen, J. Opt. Soc. Am. A **21**, 713 (2004).
7. H. L. Shen and J. H. Xin, Appl. Opt. **44**, 1969 (2005).
8. G. J. Klinker, S. A. Shafer, and T. Kanade, Int. J. Comput. Vis. **4**, 7 (1990).
9. L. Vincent and P. Soille, IEEE Trans. Pattern Anal. Mach. Intell. **13**, 583 (1991).