A STATISTICAL ERROR ANALYSIS FOR VOXEL COLORING

Musik Kwon† , *Kyoung Mu Lee*‡, *and Sang Uk Lee*†

† School of Electrical Eng., Seoul National University
‡ Dept. of Electrical and Electronics Eng., Hong-Ik University

ABSTRACT

This paper presents an error analysis for *voxel coloring* [2, 3], which is one of the well known methods to reconstruct 3D shape from 2D calibrated multiple-view images. In order to analyze the errors arising in the reconstruction process of voxel coloring algorithms, we first model several noise sources in the analytic or statistical way, and then examine the effects of each noise component on the reconstructed 3D model. Specifically, in order to analyze the statistical errors, we focus on the distribution of the image variance, which is employed as photo consistency measurement. And also, we show that how specular components induce errors in reconstructing 3D model. The results of this analysis are very useful for evaluating the statistical confidence of the reconstructed 3D model as well as finding the optimal threshold for the occupancy decision.

1. INTRODUCTION

Three-dimensional (3D) scene reconstruction from a set of images has been a fundamental problem in computer vision society. There have been a large number of works to solve it for the last couple of decades, attracting a great attention in many applications from robotics to computer graphics and virtual reality. Stereo vision is one of the traditional methods to recover the depth map [5], which however needs a notoriously difficult computation of correspondence between images. Although lots of attempts have been made to find the depth map more accurately so far, no single solution still came out to cope with various configurations of input image data. An alternative method is the shape-from-silhouette method, to recover a 3D object in a volumetric way. Typically, shape-from-silhouette method extract an object's shape from the 3D space using volume intersection based on the binary information, *i.e.* object or background, which is obtained by image segmentation. Therefore, it possibly has comparatively large ambiguities in determining the object's 3D shape. Recently, Seitz and Dyer [2] proposed an enhanced voxel-based 3D reconstruction method, referred to as voxel coloring. The main idea stems from the understanding that every color or gray value in the given images can be shape constraints because a valid reconstructed 3D scene should produce synthetic images that are same as the corresponding real input images. Thus, if the input images are given in a gray or color form rather than in a binary form, the additional photometric constraints can be enforced to improve the 3D reconstruction process. On that account, voxel coloring should present a more improved result than shape-from-silhouette method in the sense of mathematics. However, in practice, voxel coloring is usually exposed to considerable statistical noises, often yielding unsatisfactory results.

Therefore, in this paper, we address a statistical error analysis for 3D shape reconstruction by voxel coloring, which is regarded as a kind of estimation process from the statistical noisy measurements. Specifically, we consider and investigate the effects of the noises from the camera calibration, imaging process, and the spatial quantization of the volume representation. We also find how the errors finally induced by the threshold in voxel coloring. Lastly, the specular reflection is taken into account to analyze the errors caused by the specular component.

2. VOXEL COLORING ALGORITHM REVIEW

Voxel coloring divides the 3D scene space into an array of voxels first, and then traverses each voxel in an arranged order to test for the photo consistency, by which each voxel is determined to be opaque or transparent. Photo consistency has been formally defined by Seitz and Kutulakos [2]. It can be summarized as that a 3D point on a scene surface is photo-consistent between the given visible images if the color at the corresponding image pixel could have resulted from the irradiance of the 3D point. Thus, under the assumption of Lambertian reflectance, it is inferred that a valid 3D point on the surface is photo-consistent with a set of input images. Inversely, it can also be understood that photo-consistency provides the way to verifying a hypothesis about whether a point is on the 3D scene surface or not. That is, if an arbitrary 3D point is photo-consistent between multiple visible images, it is determined as being on

0-7803-7750-8/03/$17.00 ©2003 IEEE

the surface, labelled as an opaque voxel. Otherwise, it is regarded as being out of the surface, labelled as a transparent. Another contribution of voxel coloring is to present the simplified way to dealing with visibility problem. This simplification is valid if a set of cameras is placed on the same side of a plane, generating a sequence of voxel plane at increasing distance from all the cameras. Such an ordering guarantees that, when a voxel is visited to test, all possible occluding voxels have been already visited with respect to all the cameras. Thus, the visibility of current voxel is determined by the geometric structure of previously traversed voxels, which are already labelled.

3. ERROR ANALYSIS IN LAMBERTIAN SURFACE

We have a chain of two mappings in the voxel coloring. One is a mapping of a point $\mathbf{M}$ in the 3D Euclidean space onto a set $\{\mathbf{m}_k\}$ $(k = 0, 1, \cdots, n-1)$ in the 2D pixel coordinate by camera calibration $\mathbf{P}_k$, and the other is a mapping of a pixel coordinate $\mathbf{m}_k$ onto an intensity value in the color space by the image $\mathbf{I}_k$, thus we have

$$(X, Y, Z) \overset{\mathbf{P}_k}{\mapsto} \{(x_k, y_k)\} \overset{\mathbf{I}_k}{\mapsto} \{(R, G, B)_{(x_k, y_k)}\}. \quad (1)$$

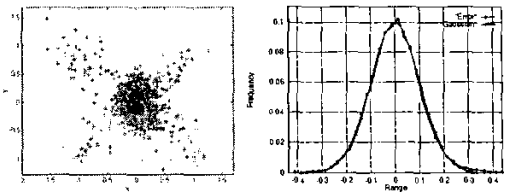
In ideal case, we have no errors in the correspondence between a 3D position (X, Y, Z) and a set of 2D position of each image. Therefore, under the assumption of Lambertian reflectance model, if all the color values in $\{(R, G, B)_{(x_k, y_k)}\}$ $(k = 0, 1, \cdots, n-1)$ are consistent, we can decide that a voxel in the position (X, Y, Z) is opaque. Practically, however, we can not avoid several sources of noise during the measurements such as camera calibration and imaging. Voxel coloring also requires volume quantization for computing. In addition, the Lambertian reflectance is far from being totally true in the real situations. Consequently, these several noise sources cause an error in color consistency.

3.1. Noise Models and Image Variance

First, let us consider the camera calibration error as well as volume quantization. Since voxel coloring first quantizes 3D space as a discrete array of voxels, it causes errors in computing the corresponding 2D image coordinate $\mathbf{m}_k$ from a given 3D point $\mathbf{M}$, as calibration error does. If we assume that calibration parameter error ΔP_k and volume quantization error ΔM in the 3D space is sufficiently small, they can be approximated as followings,

$$\begin{aligned} \widetilde{\mathbf{m}}_k &= (P_k + \Delta P_k)(M + \Delta M) \\ &\approx P_k M + P_k \Delta M + \Delta P_k M = \mathbf{m}_k + \mathbf{n}_{2d}. \end{aligned} \quad (2)$$

The volume quantization error ΔM in the 3D space follows uniform distribution in the bounded region, thus the distribution of $P_k \Delta M$ is a transformed version of uniform distribution. For camera calibration, it estimates a projection matrix from noisy image measurements of known 3D points. Therefore, errors are caused by the measurements associated with the projective relation between 3D scene and the corresponding images, which is very difficult to calculate in the analytic way. But, if we focus on the projected error $\Delta P_k M$ in pixel coordinate, the error can be modelled as a 2-dimensional additive Gaussian distribution, although it depends on the camera calibration algorithm. In order to make sure, the error distribution is plotted in Fig. 1(a). In addition, the marginal error distribution for x-axis is measured in Fig. 1(b), where a Gaussian distribution is also illustrated to show that it is a good approximation. Second, let us consider the imaging process noise, caused by optic devices. This kind of thermal noise is generally interpreted as a white Gaussian. Let $I(x, y)$ and $\widetilde{I}(x, y)$ be the irradiance and the measured intensity of a pixel at (x, y), respectively. Then we have



(a) (b)

Fig. 1. The projected camera calibration error onto image plane in (a) the 2D plane and (b) the marginal x-axis direction.

$$\widetilde{I}(x, y) = I(x, y) + I_n, \quad (3)$$

where, $I_n \sim \mathbf{N}(0, {\sigma_I}^2)$. If we denote $\widetilde{I}(\widetilde{x}_k, \widetilde{y}_k)$ for the measured intensity of projected image from 3D point, it is represented by,

$$\widetilde{I}(\widetilde{x}_k, \widetilde{y}_k) = I(x + n_{2dx}, y + n_{2dy}) + I_n. \quad (4)$$

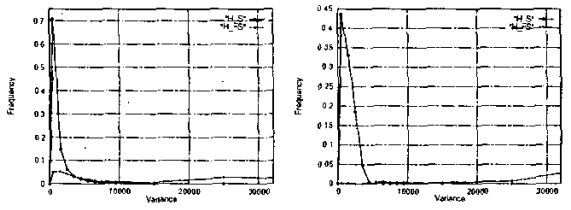
The projected error in pixel coordinate is typically so small that we can approximate it using the first order Taylor series expansion as followings,

$$I(x + n_{2dx}, y + n_{2dy}) \approx I(x, y) + \nabla_{\mathbf{n_{2d}}} I(x, y). \quad (5)$$

Under the assumption of independency between noise and image intensity, the measured variance of corresponding image intensities projected from a 3D point (X,Y,Z), denoted as $\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}(X,Y,Z)})$, is given by

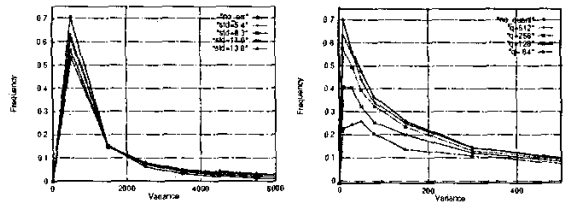
$$\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}(X,Y,Z)}) = \mathbf{var}(I_{\mathbf{P}(X,Y,Z)}) + E[\|\nabla_{\mathbf{n_{2d}}} I(x, y)\|^2] + {\sigma_\mathbf{I}}^2.$$

Notice that the measured variance is composed of three components, resulting from the desirable variance in noiseless situation, local textural gradient in the image by projection error associated with camera calibration error and volume quantization, and imaging process noise. In voxel coloring, color variance, employed as photo consistency measurement, verifies a hypothesis about whether a point is on the 3D scene surface or not, *i.e.*, $H_\mathbf{S}$ or $H_{\mathbf{S}^c}$. In the case of $H_\mathbf{S}$, the desirable variance $\mathbf{var}(I_{\mathbf{P}(X,Y,Z)})$ is zero under Lambertian reflections assumption. Thus, the variance can be rewritten in the two ways as,



(a) (b)

Fig. 2. The distribution of $\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S})$ and $\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}^c)$ is respectively plotted in Fig. 2 for (a) 'Bunny' and (b) 'Dinosaur' model.

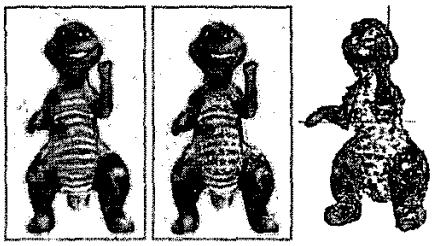


(a) (b)

Fig. 3. The changes of $\mathbf{var}\{\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}\}$ distribution by (a) camera calibration error and (b) volume quantization.

$$\begin{aligned} \mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}) &= E[\|\nabla_{\mathbf{n_{2d}}} I(x,y)\|^2] + {\sigma_\mathbf{I}}^2, \\ \mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}^c) &= \mathbf{var}(I_\mathbf{P}) + E[\|\nabla_{\mathbf{n_{2d}}} I(x,y)\|^2] + {\sigma_\mathbf{I}}^2. \end{aligned} \quad (6)$$

We can also plot the variance distributions from a set of synthetic images rendered from computer graphic models. Thus, the distributions of $\{\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S})\}$ and $\{\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}^c)\}$ for the 'Bunny' and 'Dinosaur' models are calculated and plotted in Fig. 2, respectively. As expected, $\{\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S})\}$ is concentrated on low value close to zero, while $\{\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}^c)\}$ spreads out widely. Then, let us consider how the distribution changes by the noises. For imaging noise, we can see that its variance is additive to $\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S})$ and $\mathbf{var}(\widetilde{I}_{\widetilde{\mathbf{P}}}|\mathbf{S}^c)$ from Eq.(6), but the change is very trivial to detect within the practical noise range. Fig. 3 illustrates the change of variance distribution according to the the camera calibration and volume quantization. From these figures, we can observe that, as the quantization levels become coarser or the calibration deviation gets larger, the main lobe tends to shrink and the tail gets increased, which results lots of holes. The behaviors of changes are very similar each other, because both tend to distort the position of corresponding image points in the pixel coordinate. Note that the errors due to camera calibration and quantization are closely related with local textural homogenety, which can be explained by the term $E[\|\nabla_{\mathbf{n_{2d}}} I(x,y)\|^2]$. We see that the errors grow severely where the intensities in the local region are inhomogenous around the corresponding point. Fig. 4(a) is a rendered image for Dinosaur model, where the breast and stomach are comparatively inhomogenous. In Fig. 4(b), sampled 3D points on surface having larger variance than a threshold are plotted as white dots, which are concentrated on the breast and stomach. Fig. 4(c) illustrates the reconstruction results by voxel coloring [3]. We can realize that lots of holes are caused in the locally inhomogenous region by the errors of calibration and quantization.



(a) (b) (c)

Fig. 4. Dinosaur model. (a) a rendered image, (b) the sampled 3D points on surface having larger variance than a threshold is plotted as white dots in (a) image, (b) the model reconstructed by voxel coloring.

3.2. Errors in Lambertian Surface

In voxel coloring, the errors are finally resulted from the threshold for determining either $H_\mathbf{S}$ or $H_{\mathbf{S}^c}$. The probabilistic error is given by

$$P_e = P(H_\mathbf{S}|v \in \mathbf{S}^c) + P(H_{\mathbf{S}^c}|v \in \mathbf{S}), \quad (7)$$

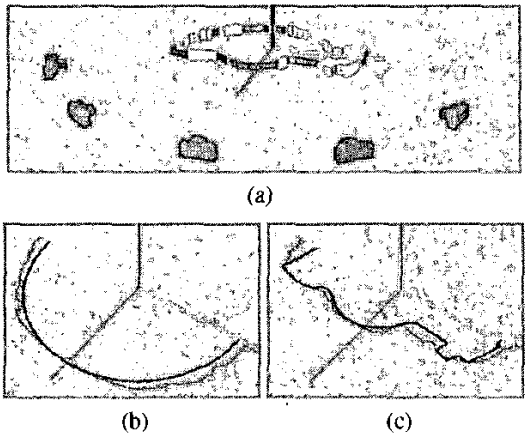
where, the first term mainly causes fattening effects on the reconstructed model, while the second gives rise to holes. Although, the optimal threshold may be determined by the maximum likelihood condition, the geometric ground truth for the calculation of the variance distributions is usually not known. Thus, instead, an optimal threshold can be found by an iterative reconstruction scheme.

3.3. Error Propagation Effects

After generating a sequence of voxel plane at increasing distance from all the cameras, voxel coloring computes visibility for a voxel from the shape composed of the previous sequence of voxel plane, and then finds its variance according to visibility. The wrong visibility tends to give a wrong estimation for current voxel. Thus, the errors are accumulated as we sweep the voxel layers. Fig. 5(a) illustrates 2-dimensional simulation of voxel coloring, and we can find that the errors are roughly increased as layers go in the Fig. 5(b) and (c).

4. SPECULAR COMPONENTS

Since the specular components generally violate the Lambertian property, the photo consistency can not be applied



(a)

(b) (c)

Fig. 5. 2-dimensional simulation (a) and results (b),(c) for error analysis of voxel coloring. In the results, the black contour is original, while the blue is reconstructed.

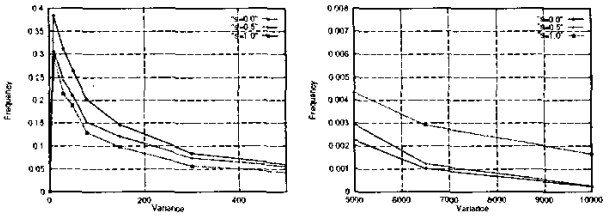
(d) (e) (f)

Fig. 6. 'Dinosaur' models are shaded as the specular components get increased. Specular factor, f_s, is set to be (a) 0.0, (b) 0.5, and (c) 1.0, respectively.

to the specular areas, resulting in significant inaccuracies in reconstruction. Thus, we also have to consider about how much influence the specular components have on the errors in voxel coloring. To understand this effect, we have rendered some test graphic models repeatedly while increasing the specular factor as illustrated in Fig. 6, and the variation of each distribution is calculated and illustrated in Fig 7. Similar to the camera calibration error, as the specular components are increased, the main lobe of $H_{\mathbf{S}}$ tends to shrink and the tail becomes more increased and longer, while no major change is conspicuous for $H_{\mathbf{S}^c}$. It is no wonder that the errors get increased as the specular components become larger. However, note that the variation is usually not so global as in the volume quantization noise case.

5. DISCUSSION AND CONCLUSION

In this paper, a statistical error analysis for voxel coloring has been presented. In detail, we have modelled several



(a) (b)

Fig. 7. (a) $\mathbf{var}(\widetilde{I}_{\mathbf{P}}|\mathbf{S})$ distributions for 'Dinosaur' models that are computed for three specular cases. (b) the tail of distribution is magnified to show that most specular one has more longer tail.

types of noises arising in voxel coloring process such as the imaging noise, camera calibration error, and the quantization noise, and then analyzed statistically how they influence on the variance distribution, respectively. Also, we have found that two different variance distributions combined with a threshold finally are directly related to the two types of error, hole or fattening effect, and the errors tends to be accumulated by the ordered sweeping process in the original voxel coloring algorithm. In addition, the effect of the specular components have been taken into consideration in a statistical way. In the future, we will extend this analysis for real scenes with ground truth. Also, for more accurate error analysis, we are going to carry out the investigation of the footprint effect in voxel coloring.

6. REFERENCES

[1] http://www.vision.caltech.edu://bouguetj/calib_doc/index.html#system.

[2] S. M. Seitz and C. R. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1067-1073, 1997.

[3] K. N. Kutulakos and S. M. Seitz, "A Theory of Shape by Space Carving," *International Journal of Computer Vision*, 38(3), pp. 199-173, 2000.

[4] W. B. Culbertson and T. Malzbender, "Generalized Voxel Coloring," *Proc. ICCV Workshop, Vision Algorithms Theory and Practice, Sringer-Verlag Lecture Notes in Computer Science 1883* , pp.100-115, Sep., 2001.

[5] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 353-363, Apr. 1993.

[6] A. Broadhurst, T.W. Drummond and R. Cipolla, "A Probabilistic Framework for Space Carving," *Proc. IEEE International Conference on Computer Vision*, pp.388-393, 2001.

[7] C. R. Dyer, "Volumetric Scene Reconstruction from Multiple Views," *Foundations of Image Understanding*, L. S. Davis, ed., Kluwer, Boston, 2001, pp.469-489.