21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

A New Metric for Measuring Image-based 3D Reconstruction

Xu Zhao, Rui Wu, Zhong Zhou*, Wei Wu
State Key Laboratory of Virtual Reality Technology and Systems,
School of Computer Science and Engineering, Beihang University, China
{*zhaoxu, wurui, zz, wuwei*}*@vrlab.buaa.edu.cn*

Abstract

We propose a novel metric to measure the image-based 3D reconstruction results without ground truth datasets. In contrast to previous metrics, our shading-based metric can not only accurately measure the reconstruction quality but also decouple from any reconstruction algorithm. Considering the uncertainty of topology, texture and soft shadow, we compute an anisotropic irradiance gradient field from multiple images to indicate the regions where reconstruction error occurs. We further apply the metric into the view planning application. Experimental results on both synthetic and real datasets illustrate the effectiveness of evaluating 3D reconstruction by our metric. The reconstruction accuracy and completeness overtop or are the same as the results of manually adding new viewpoints.

1. Introduction

Automatic reconstruction of 3D objects and environments from photographic images is important for many applications. Significant progress has been achieved since evaluation work was proposed [9, 10], but the goal is still far from the practical needs. One of the problems with image-based 3D reconstruction systems is the management of uncertainties, which arise from a variety of sources, such as the type and the position of sensors, the number and the quality of images captured, illumination conditions, the structure and the appearance of scenes to be reconstructed, and the inherent uncertainty in reconstruction algorithms.

These uncertainties result in non-optimal solution to the image-based 3D reconstruction problem, which will bring much difficulty to applications. As for input, it is hard to decide how many images are used and where are located. Hornung et al. [5] argued that it is not true that more input images would definitely lead to more accurate reconstructions. As for output, it is hard to quantitatively evaluate reconstruction results if we do not have the ground truth datasets (e.g. from 3D scanner). Obviously, it is impractical for indoor/outdoor scene modeling applications. Therefore, a measuring metric, which can guide the planning process or measure the reconstruction quality, is critical in practice.

In related work, view planning applications, including next best view planning [11, 2], image selection [5, 4] and key-frame extraction [1], concern more about metrics. After summarizing, we divide these metrics into four categories: (1) *image quality*, e.g., image entropy and signal-to-noise ratio; (2) *relationship between images*, e.g., baseline and angle between neighboring viewpoints; (3) *relationship between image and model*, e.g., overlap ratio, sampling rate and view angle; (4) *model attributes*, e.g., topology and texture. We find that parameters about the first three categories are always controllable and can be empirically determined. The fourth category, model attributes, is more related to the reconstruction process. Several metrics, like photo-consistency [5], reprojection error [1] and structure estimation uncertainty [11, 2], have been proposed. However, they are strongly coupled with the reconstruction algorithm being deployed.

Instead of using these traditional metrics, we present a more accurate metric combined with shading to measure the image-based 3D reconstruction models, and meanwhile decouple it from the reconstruction algorithms. Based on the image formation model (i.e. the rendering equation), we estimate the surface irradiance in the spherical harmonics domain from multiple images. While considering the situation of topology, texture and soft shadow, we formulate the anisotropic irradiance gradient metric to measure 3D reconstruction. Furthermore, based on this metric, we apply it into the view planning application.

The advantages of our metric include: (1) It can *automatically* indicate the reconstructed regions which are difficult to recover from images for some reasons (e.g.

*corresponding author

complex topologies, occlusion or lack of images); (2) It is *independent* of reconstruction algorithms and more *accurate* than previous metrics, which can be further used into applications such as camera planning, next best view selection and interactive model editing, etc.

2. Measuring 3D Reconstruction

Image-based modeling can be seen as the inverse process of image rendering. For a Lambertian surface, the color of the surface is essentially the projection of reflected light intensity along the direction from the point to the optical center, denoted as $I(P_c(x))$ in Fig. 1.

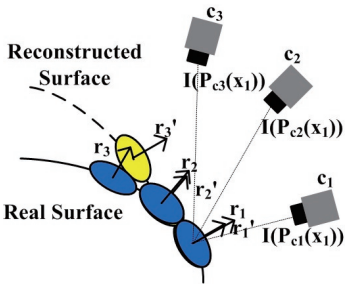

Figure 1. Measure reconstructed surface

We define the irradiance r as the reflected light intensity on 3D model surface, which depends on the surface topology, reflectance property and the consistent visible color. Here we assume the irradiance r on the object's real surface (blue patches in Fig. 1) change continuously. However, if the irradiance r' estimated on the reconstructed surface (yellow patches in Fig. 1) differs from the real irradiance, taking the r_3 patches in Fig. 1 for example, the differences between the real and the estimated irradiance become significant. Importantly, it means the irradiance gradient can indicate regions where image-based 3D reconstruction error occurs.

Note that, except the surface topology (like occlusion, sharp features etc.), the texture or the soft shadow on surface may also affect the continuity of irradiance changes, but it only happens on the shadow boundary or where texture significantly changes. This effect can be weakened by subtracting the image intensities. We will discuss the metric computation in details in the following section. The notations used in this paper are summarized in Table 1.

3. The Metric Computation

The pipeline of our computation is shown in Fig. 2. It has four steps. (1) Based on any existing MVS (Multi View Stereo) method, we create an initial 3D triangle mesh model of the object. (2) We use this model to estimate the spherical harmonic (SH) coefficients for incident radiance [8]. (3) We calculate the estimated irradiance of each vertex in accordance with the incident radiance coefficients. (4) We propose an anisotropic gradient field, the metric of the 3D reconstruction model, which indicates the reconstruction errors. Firstly, we review the image formation model in spherical harmonic domain.

Table 1. Notations

Notation	Meaning
x_i	spatial location variable for the ith vertex
$radius$	radius of the model's bounding sphere
ω_i	incident angle of the incident radiance
ω_o	outgoing angle of the reflected irradiance
Ω	domain of all incident lighting directions
$B(x)$	irradiance at the vertex x
$\mathbf{n}(x)$	surface normal of the vertex x
$Q(x)$	camera set that can see the vertex x
$Q(x_i, x_j)$	common member of $Q(x_i)$ and $Q(x_j)$
$P_c(x)$	projection of vertex x on the cth image
$I(P_c(x))$	image intensity of $P_c(x)$
$G(x)$	anisotropic gradient of the vertex x
$N(x)$	neighbor set of the vertex x
$d(p_i, p_j)$	distance between location p_i and p_j

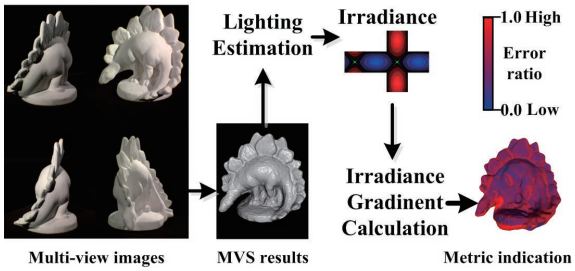

Figure 2. Metric computation overview

The Rendering Equation

Assuming all objects in the scene are non-emitters and the light sources are infinitely distant, the rendering equation [6] can be defined as:

$$B(x, \omega_o) = \int_{\Omega} L(\omega_i) V(x, \omega_i) \rho(\omega_i, \omega_o) max(\omega_i \cdot n(x), 0) d\omega_i \quad (1)$$

where $B(x, \omega_o)$ is the irradiance at vertex x to the direction ω_o and $L(\omega_i)$ represents the incident lighting. $V(x, \omega_i)$ is a binary visibility function, and $\rho(\omega_i, \omega_o)$ is the BRDF (Bidirectional Reflectance Distribution Function) of the surface. For convenience, we simply scale the incident radiance by the albedo, letting

$L_a(\omega_i) = \rho L(\omega_i)$.

Lighting Estimation

Based on the MVS result, we estimate the SH coefficients for the incident radiance. We define $T(x, \omega_i) = V((x, \omega_i) max(\omega_i \cdot n(x), 0)$. Then we project $T(x, \omega_i)$ and $L_a(\omega_i)$ into SH domain, and get the SH coefficients l_k, t_k. According to the orthogonality of SH basis functions, the rendering equation becomes

$$B(x, \omega_o) = \int_\Omega L_a(\omega) T(x, \omega_i) d\omega_i = \sum_{k=1}^{n^2} l_k t_k \quad (2)$$

Then, we calculate the coefficients of incident radiance l_k by minimizing

$$\hat{l} = \arg\min_l \sum_l \sum_{c \in Qi} |\sum_{k=1}^{n^2} l_k t_k - I(P_c(x_i))| \quad (3)$$

Irradiance Estimation

Given the estimated geometry and incident radiance, this step is to calculate irradiance based on the rendering equation. If we define $L_v(\omega) = L_a(\omega_i) V(x, \omega_i)$, the rendering equation can be represented as:

$$B(x, \omega_o) = \int_\Omega L_v(\omega) max(\omega_i \cdot n(x), 0) d\omega_i \quad (4)$$

Based on the Funk-Hecke-Theorem, the circular symmetric function $max(\omega_i \cdot n(x), 0)$, can be parameterized into a set of scalars $\{a_k\}$. Then we get

$$B(x) = \sum_{k=0}^{n-1} \sum_{m=-k}^{k} \sqrt{\frac{4\pi}{2k+1}} a_k g_{km} Y_{km} \quad (5)$$

where g_{km} is the SH coefficients of the function $L_v(\omega)$, and Y_{km} is the SH basis function.

Anisotropic Gradient Field Calculation

Shape-from-shading methods can recover fine details from shading variations, for they compute per-pixel surface orientation instead of sparse depth. Here we choose the irradiance information to accurately measure 3D reconstruction quality. Considering the uncertainty of texture and soft shadow, we use the differences between the measured irradiance (i.e. the image intensity) and the estimated irradiance to formulate the metric. Additionally, to preserve fine details, we use an anisotropic gradient of irradiance as a substitute.

For each vertex, the gradient is defined as:

$$G(x_i) = \sum_{x_j \in N(x_i)} \sum_{c \in Q(x_i, x_j)} w(x_i, x_j) \cdot (d_I(x_i, x_j) - d_c(x_i, x_j))^2 \quad (6)$$

where $d_I(x_i, x_j)$ and $d_c(x_i, x_j)$ are the estimated irradiance difference and measured image intensity difference of x_i,x_j, respectively. Namely,

$$\begin{aligned} d_I(x_i, x_j) &= B(x_i) - B(x_j) \\ d_c(x_i, x_j) &= I(P_c(x_i)) - I(P_c(x_j)) \end{aligned} \quad (7)$$

The weight function is negative correlation to the distance between x_i and x_j and defined as:

$$w(x_i, x_j) = radius / d(x_i, x_j) \quad (8)$$

Note that if there is no camera can see the vertex x_i and its neighbor x_j at the same time, it means that the mesh model region of x_i, x_j is estimated by the surface reconstruction method, and there are not sufficient views for this region. Thus, we set the gradient to infinity, which indicates the incomplete reconstruction areas.

4. Application: View Planning for MVS

One of the fields that our metric could be applied into is the view planning for Multi-view Stereo reconstruction. Based on our metric, we propose a new view planning algorithm, as shown in Fig. 3.

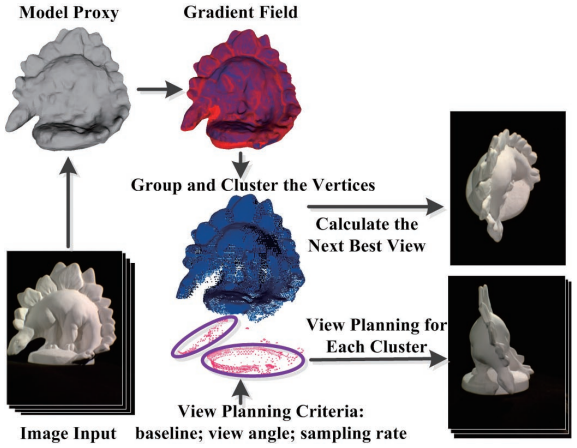

Figure 3. Our metric based view planning

To formulate the planning objective function (Eq. 9), we refer to some empirical criteria, including baseline b, sampling rate r, view angle $angle$. We divide the vertices into two groups: (1) the vertices whose gradient is set to infinity; (2) other vertices. And we perform different planning strategies on them. The view planning function is minimized with the Active-Set algorithm. Then we add a set of views for each cluster of group (1), and a next best view for group (2).

$$\min_{pos,ori} \sum_i G(x_i) \cdot \frac{|b(pos, pos_c) - b_0|}{angle(ori, v(pos, x_i))} \cdot r(x_i, pos) \quad (9)$$

For implementation, we use Furukawa's PMVS (Patch-based Multi-View Stereo) method [3] and Poisson surface [7] to generate the initial MVS models, calculate the metric and guide the view planning. We set the band of spherical harmonics to 3 as a compromise of accuracy against efficiency. The angle between the neighbor views is set to 20°.

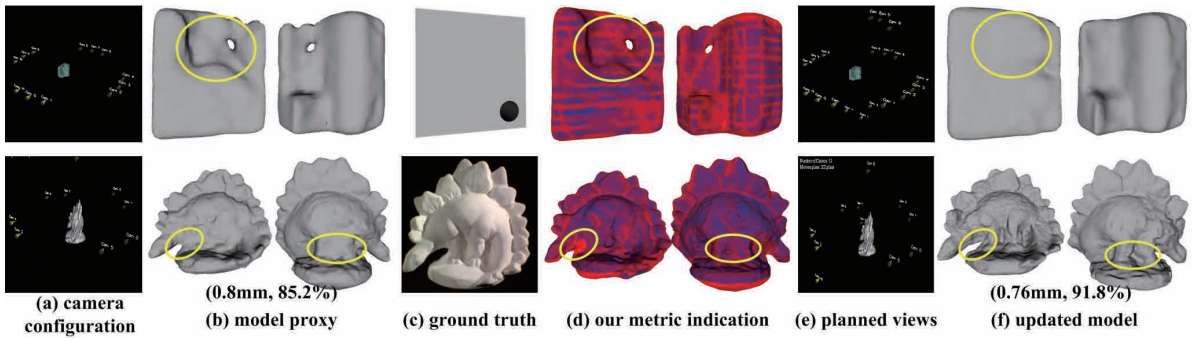

Figure 4. 3D reconstruction evaluation on both synthetic data (upper) and real images (lower)

In the synthetic experiment, due to insufficient image input, some poorly reconstructed regions (e.g. parts marked in yellow circle, the top of model, and the hole parts) are revealed by high gradient (Fig. 4(d)). After view planning, three views from top and one extra view for the body are added. It is obvious that in the updated model these regions gain a much better appearance.

Experiments on real images are performed using the Middlebury dino dataset. The metric sheds light on the reconstruction errors and topologically important surface regions (Fig. 4(d)), including the neck parts with obvious reconstruction error and the leg parts which need to be refined. The updated model shows more details and an accurate reconstruction of the neck (Fig. 4(f)). According to the Middlebury Evaluation system, the updated model is better in completeness (91.8% against 85.2%) and accuracy (0.76mm against 0.8mm). Furthermore, we compare the results updated by our view planning algorithm with the method of manually adding new viewpoints (according to the experts' experience). Our algorithm shows better performance in completeness (91.8% against 90.8%), while as good capability in accuracy (0.76mm against 0.70mm).

5. Conclusion

This paper presents a new metric based on the anisotropic irradiance gradient to measure 3D reconstruction. The metric can automatically indicate the regions where any image-based 3D reconstruction algorithms cannot recover well. Besides, we successfully apply it to the view planning application. The main limitation is the current metric computation is seriously constrained by the initial reconstructed surface meshes. We would like to study the irradiance-consistent based surface reconstruction method to relax this dependency.

Acknowledgements

This work is supported by the National 863 Program of China under Grant No.2012AA011803, the Natural Science Foundation of China under Grant No.61170188, the National 973 Program of China under Grant No.2009CB320805, and Fundamental Research Funds for the Central Universities of China.

References

[1] M. T. Ahmed, M. N. Dailey, J. L. Landabaso, and N. Herrero. Robust key frame extraction for 3d reconstruction from video streams. In *VISAPP*, pages 231–236, Angers, France, 2010.

[2] E. Dunn and J. M. Frahm. Next best view planning for active model improvement. In *BMVC*, pages 1–11, London, UK, 2009.

[3] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *PAMI*, 32(8):1362 – 1376, 2010.

[4] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *ICCV*, Rio de Janeiro, Brazil, 2007.

[5] A. Hornung, B. Zeng, and L. Kobbelt. Image selection for improved multi-view stereo. In *CVPR*, pages 1–8, Anchorage, AK, 2008.

[6] J. T. Kajiya. The rendering equation. In *SIGGRAPH*, pages 143–150, New York, USA, 1986.

[7] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *SGP*, pages 61–70, Aire-la-Ville, Switzerland, 2006.

[8] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. *JOSA*, 18(10):2448–2459, 2001.

[9] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, pages 519–528, Washington, DC, USA, 2006.

[10] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, pages 1–8, Anchorage, AK, 2008.

[11] S. Wenhardt, B. Deutsch, E. Angelopoulou, and H. Niemann. Active visual object reconstruction using d-, e-, and t-optimal next best views. In *CVPR*, pages 1 –7, Minnesota, USA, 2007.