ERROR ANALYSIS FOR IMAGE-BASED RENDERING WITH DEPTH INFORMATION

Ha T. Nguyen and Minh N. Do

Department of Electrical and Computer Engineering,
University of Illinois at Urbana Champaign.
{hanguyen, minhdo}@uiuc.edu.

ABSTRACT

We propose a novel approach to analyze the rendering error of image-based rendering (IBR) algorithms with depth information. We do not use the assumption of band-limitedness as existing approaches. Instead, we use the framework of the Propagation Algorithm that allows to rigorously analyze the rendering error via the framework of nonuniform interpolation. In this framework, using the depths, we propagate all the intensity information to the virtual cameras, and by doing so, turning the IBR problem into a nonuniform interpolation problem at the virtual image planes. The proposed approach then can systematically analyze the rendering quality for different interpolation methods, including commonly used linear interpolation. We can furthermore analyze the effect of depth estimation error on the rendering quality.

1. INTRODUCTION

Image-based rendering (IBR) is an emerging technology that has been developed as an alternative to traditional *model-based techniques* for image synthesis. IBR synthesizes novel (or virtual) images, as taken by virtual cameras at arbitrary viewpoints, using a set of acquired images. With the advantages of being photorealism and having low complexity over model-based techniques, IBR has many potential applications such as remote reality and telepresence [8].

The depth information has been used in literature as a crucial information in addition to the intensity information. Most of the existing IBR algorithms require the depth information, either explicitly or implicitly in form of correspondences. In [8], Shum et al. discuss levels of prior knowledge of the scene geometry needed in existing IBR algorithms.

Although many algorithms have been proposed to render virtual images, to our knowledge, no method has been proposed for the error analysis of IBR algorithms. In [1, 9], the plenoptic function is analyzed in the frequency domain, by assuming that the plenoptic function is band-limited, and this analysis can lead to an error analysis. However, the assumption of band-limitedness is a global assumption and it does not necessary hold in practice. Do et al. show that, in general, the plenoptic function is not bandlimited unless the surfaces are flat [4]. Furthermore, in contrary to the extended use of the depth information in IBR algorithms, the depth information is rarely exploited in the analysis of the plenoptic function.

In this paper, we propose two novel aspects to tackle the problem of error analysis for IBR algorithms. First, we use a local instead of a global approach by adopting the framework of nonuniform interpolation. This framework includes some techniques commonly used in practice such as B-splines [3] (with order 1 is in fact linear interpolation). As a consequence, the error analysis of our method will be faithful to the actual error encountered in practice. Second, the depth information is considered available at all image pixels, either thanks to range cameras or structure from motion techniques. Our analysis is able to measure the effect of depth estimation error on the rendering quality.

This paper is organized as follows. We set up the problem and briefly describe the Propagation Algorithm in Section 2. Section 3 introduces a bound for the interpolation error depending on the intersample gaps and jitters. Then in Section 4, we propose bounds for the intersample gaps and jitters. In Section 5, we derive an error bound for the rendering quality. Section 6 shows an experiment. Finally, we conclude the paper and give discussions in Section 7.

2. PROBLEM STATEMENT

2.1. The scene model

We study the 2D light field model [5] as illustrated in Fig. 1. The cameras are pinhole cameras whose perspective centers lie on the line $d = 0$ with parameter u. The camera focal lines lie at $d = 1$ with parameter v. The field of views of cameras are supposed to be large enough to cover the scene.

We consider the scene model that consists of a *non-self-occluded* surface characterized by the depth function $d(u) \in (d_{min}, d_{max})$ for all $u \in (u_{min} - \epsilon, u_{max} + \epsilon), \epsilon > 0$[1], and an texture function "painted" on the surface. This model represents a micro-scale analysis of the plenoptic function, where

This work was supported by the National Science Foundation under Grant ITR-0312432.

[1]This is to ensure that the cameras do not lie on tangents of the depth function, which is a special case of self-occlusion.

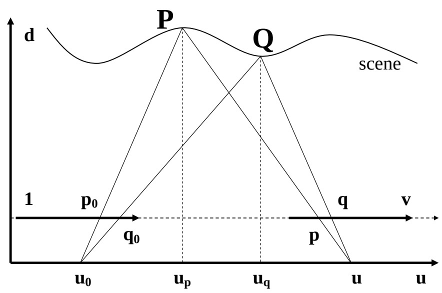

Fig. 1. The scene model. Coordinates u, v and d specify the camera position, pixel position, and the depth, respectively.

locally only one object with its surface is visible. The scene is supposed to be Lambertian.

We consider a set of actual cameras whose centers are in the interval $[u_{min}, u_{max}]$. Suppose all the cameras have the same resolution Δ characterized by the (equal) space between consecutive pixels on the image plane v. We want to investigate the effect of the number of camera N_c, the camera resolution Δ, and the scene geometry (under a condition number σ defined later) in the rendering quality.

2.2. Rendering using the Propagation Algorithm

In this paper, we investigate the analysis of the rendering error for the Propagation Algorithm [6] because it allows us to rigorously analyze the rendering error via the framework of nonuniform interpolation. Nevertheless, the analysis of the paper is applicable to other algorithms for IBR with depth information as well; e.g. [2, 7]. The main ideas of the Propagation Algorithm are as follows.

Information Propagation. Using the depth information, we propagate all the available information to the virtual camera image plane. For example, in Fig. 1, the intensity and depth information at actual pixel p_0 of actual camera u_0 are propagated to point p at virtual camera u. Note that p may not be at pixel position.

Occlusion Removal. We remove all the points in whose neighborhood there is another point with noticeably smaller depth; these points are likely occluded at the virtual camera. This step is crucial when we consider occluded scenes. However, in this paper, this step is irrelevant because the scene is supposed to be non-self-occluded.

Intensity Interpolation. In the virtual image plane, we interpolate the remaining points based on the framework of nonuniform interpolation [3].

3. ERROR ANALYSIS FOR THE INTENSITY INTERPOLATION STEP

The rendered image is obtained after the Intensity Interpolation step using the points propagated from the actual cameras. In this section, we introduce a bound for the rendering error for the case of linear interpolation for its common use in practice. Note that the results can be generalized [3].

Proposition 1. *We consider a function $f(x)$ in $C^2[\mathbb{R}]$. Let $\widehat{f}(x)$ is the linear interpolation of $f(x)$ from $f(a+\gamma_a), f(b+\gamma_b)$ with γ_a, γ_b are sample jitters of a, b. For $x \in [a,b]$, the interpolation error of $\widehat{f}(x)$ is bounded by:*

$$|\widehat{f}(x) - f(x)| \leq \frac{(b-a)^2}{8}\|f''\|_\infty + \max\{|\gamma_a|, |\gamma_b|\}\|f'\|_\infty. \tag{1}$$

Proof. See Appendix 8.1. □

Given the intensity function $f(x)$ at the virtual image plane that we want to reconstruct at pixel locations, the linear interpolation error is bounded by a function of the intersample gap $(b-a)$ and the sample jitters γ_a, γ_b. We will see in Section 4 how the sample gaps and sample jitters can be bounded.

4. BOUNDS FOR GAPS AND JITTERS AT THE INFORMATION PROPAGATION STEP

In the Information Propagation step, we propagate all available information from the actual cameras to the virtual camera. This turns the IBR problem into a nonuniform interpolation problem at the virtual image plane. In Section 4.1 we introduce a bound for the gaps between propagated points at the virtual image plane. In Section 4.2, we will give bounds for the sample jitters caused by a wrong estimate of the depth.

The notation for this section follows Fig. 1, with u, u_0 are cameras; P, Q are points in the scene whose coordinates are $(u_p, d(u_p))$ and $(u_q, d(u_q))$; p, p_0 are images of P at cameras u, u_0; and q, q_0 are images of Q at cameras u, u_0.

4.1. Bound for sample gaps

In this subsection, we propose a bound for the gaps in the virtual image plane of points propagated from an actual camera. The result will allow us to analyze the effect of the camera resolution on the final rendering quality.

Proposition 2. *Let p_0, q_0 are two image points of an actual camera u_0, and p, q are their propagated points at a virtual camera u. There exist a constant $\sigma > 0$ such that:*

$$|p - q| \leq \sigma|p_0 - q_0|. \tag{2}$$

Proof. See Appendix 8.2 □

For an actual camera u_0, if we propagate all the actual pixels to a virtual camera u, we will have a set of points $\mathcal{X} = \{x_n\}_{n=1}^N$, where $x_{n+1} \leq x_n$, on the virtual image plane. Thanks to proposition 2, we have a bound for $(x_{n+1} - x_n)$:

$$0 \leq x_{n+1} - x_n \leq \sigma\Delta, \tag{3}$$

The gap bound in (3) depends both on the camera resolution Δ and the number σ. The number σ can be considered as the condition number that combines the geometrical structure of the scene (through the depth function $d(x)$) and the position of the (actual and virtual) cameras. Decreasing σ will help to obtain a better error bound. This fact may have an impact of where to put the actual cameras around the scene given a virtual camera position.

4.2. Bound for sample jitters

In Section 4.1, we propose a bound for the gaps between propagated points. In order to propagate the actual pixels to the virtual image plane, we need the depth information. In practice, the depth information is subject to estimation error. An error of the depth at an actual pixel will result a jitter in the virtual image plane. In this subsection, we will give a bound for the jitters.

Proposition 3. *Let p_0 is a pixel at actual camera u_0 with depth $d(u_p)$. Suppose p and $\widehat{p}$ are the position of the point propagated from p_0 in the image plane of the virtual camera u in the case we have the right depth $d(u_p)$ and the wrong depth $d(u_p) + \epsilon$, respectively. The jitter $\gamma_p = \widehat{p} - p$ will satisfy:*

$$|\gamma_p| \leq \frac{|u - u_0| \cdot |\epsilon|}{d_{min}(d_{min} - |\epsilon|)}. \qquad (4)$$

Proof. See Appendix 8.3. □

As shown in (4), the the sample jitters can be bounded using a bound of the depth estimation error ϵ.

5. ERROR ANALYSIS FOR IBR

In this section, let us consider a virtual camera with the intensity function $f(x)$. We suppose that $f(x)$ has second derivative on the virtual image plane. Note that this assumption is looser than the assumption of bandlimitedness, and our analysis can be extended for more general classes of $f(x)$. The following theorem gives an error bound for a virtual pixel.

Theorem 1. *Given an actual camera with resolution Δ, depth error bound ϵ, the intensity error e of a* virtual pixel *in the case of linear interpolation can be bounded by:*

$$|e| \leq \frac{\sigma^2 \Delta^2}{8} \|f''\|_\infty + \frac{|u - u_0| \cdot \epsilon}{d_{min}(d_{min} - \epsilon)} \cdot \|f'\|_\infty. \qquad (5)$$

Proof. Using Propositions 1, 2 and 3. □

Remark 1. *The bound for the intensity error e in Theorem 1 is for the case of one actual camera only. This bound depends on the camera resolution Δ, the geometry of the scene σ, and the depth error bound ϵ. In fact, if we have more than one cameras, the sample gaps $(x_{n+1} - x_n)$ could be better bounded and depends on N_c. Analyzing the number of actual cameras needed is the problem of plenoptic sampling, which we also intend to investigate in the direction of this paper.*

6. EXPERIMENT

We run our experiment for the case where the actual camera is at $u_0 = 3.14$ and the virtual camera is at $u = 5$. We consider the scene of constant depth $d = 10$ with depth error $|\epsilon| \leq 0.2$ in the interval $[u_{min}, u_{max}] = [0, 10]$. The intensity function "painted" on the scene is $I(t) = \sin(t)$, hence the intensity function at the virtual camera is $f(x) = \sin(10x)$. The virtual image is rendered using the Propagation Algorithm [6].

Fig 2 shows the rendering error of virtual pixels (line) compared to the error bound derived in Theorem 1 (dash). We can see that the error bound gives a good indication of the rendering error.

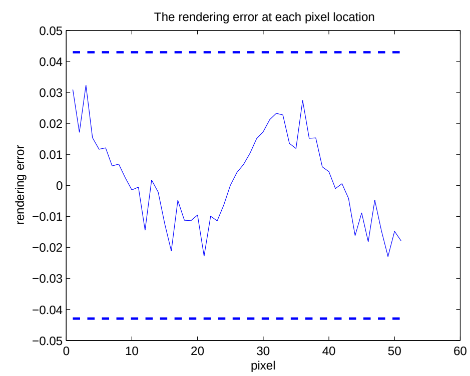

Fig. 2. The rendering error (line) and the theoretical bound (dash). The bound is computed using Theorem 1.

7. CONCLUSIONS AND DISCUSSIONS

In this paper, we propose a new approach to analyze the rendering error of IBR algorithms with depth information. We do not use the assumption of band-limitedness as existing approaches. Instead, we use the framework of the Propagation Algorithm that allows to rigorously analyze the rendering error via nonuniform interpolation. The rendering quality then can be analyzed for different interpolation methods, including piecewise constant and linear interpolation. We can furthermore analyze the effect of depth estimation error on the rendering quality.

The result in Theorem 1 is a special case where we consider only one actual camera. We plan to tighten the bound for the sample gaps $(x_{n+1} - x_n)$ in the case where several actual cameras are available. We predict that the problem of plenoptic sampling can be approached in this direction.

The number σ in (2) can be interpreted as the condition number combining the geometrical structure of the scene and the positions of the (actual and virtual) cameras. We plan to further investigate this number with expectation that it helps to put the actual cameras at the optimal locations given the scene and a virtual camera position.

Separating the effect of the texture and of the scene geometry on the final rendering quality is also a possible issue. Let us consider the components $f''(x)$ and $f'(x)$ in Theorem 1; where $f(x)$ is in fact the composite function of the texture function and the point-to-pixel mapping function (the function $g_u(x)$ in Appendix 8.1). Hence, taking derivatives of $f(x)$ can help analyzing the impact of the texture function and the scene geometry on the rendering quality. This is also a future research.

8. APPENDIX

8.1. Proof of Proposition 1

We first use the Taylor expansion to obtain:

$$f(b+\gamma_b) = f(b) + \gamma_b f'(\alpha_b),$$
$$f(b) = f(x) + (b-x)f'(x) + \frac{1}{2}(b-x)^2 f''(\beta_b).$$

for some appropriate α_b, β_b. Hence:

$$f(b+\gamma_b) = f(x) + (b-x)f'(x) + \frac{1}{2}(b-x)^2 f''(\beta_b) + \gamma_b f'(\alpha_b). \quad (6)$$

Similarly, we can obtain the following equation for a:

$$f(a+\gamma_a) = f(x) + (a-x)f'(x) + \frac{1}{2}(a-x)^2 f''(\beta_a) + \gamma_a f'(\alpha_a). \quad (7)$$

Using (6) and (7) we get:

$$\widetilde{f}(x) - f(x) = \frac{b-x}{b-a} f(a+\gamma_a) + \frac{x-a}{b-a} f(b+\gamma_b) - f(x)$$
$$= \frac{1}{2}\frac{(b-x)(x-a)^2}{b-a} f''(\beta_a) + \frac{b-x}{b-a}\gamma_a f'(\alpha_a) +$$
$$+ \frac{1}{2}\frac{(b-x)^2(x-a)}{b-a} f''(\beta_b) + \frac{x-a}{b-a}\gamma_b f'(\alpha_b).$$

Taking the absolute value of both sides:

$$|\widetilde{f}(x) - f(x)| \le \frac{1}{2}(b-x)(x-a)\|f''\|_\infty + \max\{|\gamma_a|, |\gamma_b|\}\|f'\|_\infty$$
$$\le \frac{1}{8}(b-a)^2\|f''\|_\infty + \max\{|\gamma_a|, |\gamma_b|\}\|f'\|_\infty.$$

8.2. Proof of Proposition 2

Consider the virtual camera u. We define a point-to-pixel mapping function $g_u(x)$ that maps each point $(x, d(x))$ in the scene to its image point at the camera u:

$$g_u(x) = \frac{x-u}{d(x)}.$$

The derivative of $g'_u(x)$ can be computed as:

$$g'_u(x) = \frac{d(x) - d'(x)(x-u)}{d(x)^2}.$$

We define constants A, B for camera u as follows:

$$A = \inf_{[u_{min}, u_{max}]} g'_u(x),$$
$$B = \sup_{[u_{min}, u_{max}]} g'_u(x).$$

Lemma 1. *For all camera $u \in [u_{min}, u_{max}]$ we have:*

$$0 < A < B < +\infty.$$

Proof. The proof is based on the assumption of non-self-occlusion. More details will be available in our coming technical reports. □

Let p, q be images of points P, Q, then we have $p = g_u(u_p)$ and $q = g_u(u_q)$. Hence, there exists $\theta \in [u_p, u_q]$ such that:

$$p - q = g_u(u_p) - g_u(u_q) = (u_p - u_q) g'_u(\theta).$$

As $0 < A \le g'_u(\theta) \le B < +\infty$, the following holds:

$$A|u_p - u_q| \le |p-q| \le B|u_p - u_q|.$$

The same for the actual camera u_0:

$$A_0|u_p - u_q| \le |p_0 - q_0| \le B_0|u_p - u_q|.$$

The number $\sigma = B/A_0$ satisfies Proposition 2.

8.3. Proof of Proposition 3

Consider a point P in the scene. By simple derivation we can get:

$$p - p_0 = -\frac{u-u_0}{d(u_p)}.$$
$$\widetilde{p} - p_0 = -\frac{u-u_0}{d(u_p)+\epsilon}.$$

Hence, the jitter in the virtual image plane will be:

$$\gamma_p = \widetilde{p} - p = (u-u_0)\left(\frac{1}{d(u_p)} - \frac{1}{d(u_p)+\epsilon}\right).$$
$$= \frac{\epsilon(u-u_0)}{d(u_p)(d(u_p)+\epsilon)}.$$

Because $d(x) \ge d_{min}$, the proposition is proven.

9. REFERENCES

[1] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proc. of SIGGRAPH*, 2000, pp. 307–318.

[2] S. E. Chen and L. Williams, "View interpolation for image synthesis," in *Proc. of SIGGRAPH*, 1995, pp. 29–38.

[3] C. de Boor, *A practical guide to splines*. Springer-Verlag, 1978.

[4] M. N. Do, D. Marchand-Maillet, and M. Vetterli, "On the bandlimitedness of the plenoptic function," in *Proc. IEEE Int. Conf. on Image Proc.*, 2005.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. of SIGGRAPH*, 1996, pp. 31–40.

[6] H. T. Nguyen and M. N. Do, "Image-based rendering with depth information using the propagation algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Philadelphia, 2005.

[7] S. M. Seitz and C. M. Dyer, "View morphing," in *Proc. of SIGGRAPH*, 1996, pp. 21–30.

[8] H. Y. Shum, S. B. Kang, and S. C. Chan, "Survey of image-based representations and compression techniques," *IEEE Trans. Circ. and Syst. for Video Tech.*, vol. 13, pp. 1020–1037, 2003.

[9] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Trans. Circ. and Syst. for Video Tech.*, 2003.