2015

Optimization of the number of rays in interpolation for light field based free viewpoint systems

Hooman Shidanshidi
University of Wollongong, hooman@uow.edu.au

Farzad Safaei
University of Wollongong, farzad@uow.edu.au

Wanqing Li
University of Wollongong, wanqing@uow.edu.au

Optimization of the number of rays in interpolation for light field based free viewpoint systems

Abstract

Light field (LF) rendering is widely used in free viewpoint video systems (FVV). Different methods have been proposed to employ depth maps to improve the rendering quality. However, estimation of depth is often error-prone. In this paper, a new method based on the concept of effective sampling density (ESD) is proposed for evaluating the depth-based LF rendering algorithms at different levels of errors in the depth estimation. In addition, for a given rendering quality, we provide an estimation of number of rays required in the interpolation algorithm to compensate for the adverse effect caused by errors in depth maps. The proposed method is particularly useful in designing a rendering algorithm with inaccurate knowledge of depth to achieve the required rendering quality. Both the theoretical study and numerical simulations have verified the efficacy of the proposed method.

OPTIMIZATION OF THE NUMBER OF RAYS IN INTERPOLATION FOR LIGHT FIELD BASED FREE VIEWPOINT SYSTEMS

Hooman Shidanshidi, Farzad Safaei, Wanqing Li

ICT Research Institute, University of Wollongong, Australia
hooman@uow.edu.au, farzad@uow.edu, wanqing@uow.edu.au

ABSTRACT

Light field (LF) rendering is widely used in free viewpoint video systems (FVV). Different methods have been proposed to employ depth maps to improve the rendering quality. However, estimation of depth is often error-prone. In this paper, a new method based on the concept of effective sampling density (ESD) is proposed for evaluating the depth-based LF rendering algorithms at different levels of errors in the depth estimation. In addition, for a given rendering quality, we provide an estimation of number of rays required in the interpolation algorithm to compensate for the adverse effect caused by errors in depth maps. The proposed method is particularly useful in designing a rendering algorithm with inaccurate knowledge of depth to achieve the required rendering quality. Both the theoretical study and numerical simulations have verified the efficacy of the proposed method.

Index Terms— Light Field Rendering, Free Viewpoint Video, Lossy Reconstruction, Ray Interpolation

1. INTRODUCTION

A free viewpoint video (FVV) system aims to provide users with the ability to control their viewpoint in real-time. Light field (LF) is a simplified four dimensional plenoptic function [1] first introduced by Levoy and Hanrahan [2] and Gortler et al [3] (as Lumigraph) in mid 1990s and is one of the promising approaches to realize FVV. An LF-based FVV system consists of three main components: *LF acquisition*, *LF rendering* and *LF compression/transmission*. The compression and transmission component is outside the scope of this paper and will not be considered further.

LF acquisition (i.e., plenoptic signal sampling) is concerned with sampling a subset of rays from the scene using a number of cameras at a given sampling density (SD). SD at a given location can be defined as the number of samples acquired per unit area of the convex hull of the surface of the scene at that location. A simple light field acquisition model uses a camera grid and specifies the rays by their intersection points with two parallel planes/slabs. The limitations of this model have been addressed by more complicated ray parameterizations such as two-sphere (2SP) and sphere-plane parameterizations (SPP) [4]. However, the the two-plane model provides an effective way for system analysis as the results can be generalized to the other models.

LF rendering (i.e., plenoptic signal reconstruction) aims to generate any user-selected view by synthesizing the unknown rays via interpolation of acquired rays. Many rendering methods have been developed so far. Some of them assume that the light field has been sampled sufficiently and employ a simple view interpolation process. Recent studies [5-10] have shown that implicit or explicit use of geometric information, such as a depth map of the scene, can significantly improve the rendering quality since in most practical cases, the light field is highly under-sampled. Typical methods include layered light field [5], surface light field [6], scam light field [7], pop-up light field [8], all-in-focused light field [9], and dynamic reparameterized light field [10].

In general, the LF rendering component consists of two processes: the *ray selection process* is responsible for choosing a subset of captured rays, purported to be in the vicinity of the unknown ray; and the *interpolation process* will estimate the unknown ray from these selected rays. The *ray selection process*, which is the focus of this paper, is often prone to error. For example, imperfect knowledge of depth may cause this process to miss some neighboring rays and choose others that are indeed sub-optimal for interpolation. Also, constraints on computational load (imposed due to, say, real-time rendering requirements) may necessitate this process to select only a subset of neighboring rays, less than what is available. In both cases, there is some loss of information and the output of this process represents an *effective sampling density* (ESD) which is *lower* than the SD obtained by the acquisition component. ESD is defined as the number of rays per unit area of the scene that have been captured by *acquisition* component and chosen by *ray selection process* to be employed in the rendering. Not surprisingly, it has been shown that ESD is the true indicator of output quality, *not* SD ─ except that ESD is fundamentally bound by SD [11, 12]. In addition, ESD provides an analytically tractable way for evaluating the influence of the imperfections of *both* acquisition and rendering components on output quality. The focus of this paper is to use the concept of ESD and quantify the tradeoff between ray selection, depth error and rendering quality.

Despite the extensive research in the use of depth information in LF rendering, evaluation of the efficacy of rendering methods has been restricted to subjective visual comparison. A typical approach to compensating for the errors in depth maps is to increase the number of cameras of acquisition component [13-16], that is, to increase the SD. In contrast, this paper investigates how the adverse effect caused by errors in depth maps can be compensated for by employing optimal number of rays in the *ray selection process* for a fixed acquisition camera grid. The current LF rendering methods often assume a linear interpolation over 4 rays in the camera plane or 16 rays in both camera and image planes in the rendering process, despite the fact that more rays may be available. We will develop an optimization model to obtain the optimum number of rays for a given output quality and depth map estimation error level. For the remainder of the paper, a given

acquisition system is assumed and also quality degradation as a result of compression/transmission is ignored.

The main contributions of this paper are as follows:

- Proposing an analytical model based on ESD to study the impact of depth estimation errors on *ray selection process* and rendering quality.
- Demonstrating that the degradation of rendering quality caused by the errors in depth estimation can, to some extent, be quantified by ESD and compensated for by selecting more rays during interpolation.
- Deriving a closed-form expression to calculate the optimal number of rays required to compensate for errors in depth map in order to meet the specified rendering quality and computational efficiency.

In addition, numerical simulations were conducted to verify the proposed model and promising results have been obtained.

1.1. Related Work

The efficacy of both acquisition and rendering components directly affect the quality of an FVV system. FVV quality assessment has been mainly based on subjective evaluation and comparison [17-20] and is usually limited to case-based studies. For instance, pixel fidelity indicators [20] or human visual system (HVS) metrics [21-23] with respect to ground-truth images [24, 25] or no-reference metric [26] were reported. However, none of these methods address the impact of depth map estimation error on the video quality and neither the exact effect of acquisition and rendering.

To analyze the effect of acquisition component on output quality, several studies [13-16] have been reported on the minimum required sampling rate of an ideal LF acquisition, i.e., minimum density of the camera grid by assuming a perfect signal reconstruction. It has also been shown that the adverse effect caused by the depth errors can be to some extent compensated for by increasing the number of cameras used in acquisition, which may not be affordable in practice.

On the other hand, the effect of rendering component on video quality has been reported in a few studies such as [11, 12, 27, 28] by analytical objective assessment of FVV video quality. Among these proposed models, [11, 12] is focused on light field quality assessment based on ESD which is adopted as the base of analysis in this paper.

Our observation is that both the rendering quality and tolerance to the errors in depth can be improved significantly by increasing ESD which can be achieved by employing more rays during the *ray selection process* without necessarily increasing the number of cameras.

2. THE PROPOSED MODEL

2.1. Overview of ESD

Let Θ be the set of all known rays captured by cameras, that is, the samples of the scene obtained during the LF acquisition phase. A rendering method uses a subset ω of rays from Θ, purported to be surrounding an unknown ray r, and interpolate them to estimate r. Assume that r intersects with the scene at point p at depth d. A is an imaginary convex hull area around p which intersects with all the rays in ω at depth d. The size of A would depend on the choice of ω made by the rendering method.

There are usually more rays from Θ passing through A but are not selected by the rendering process. However, using them effectively can potentially enhance the interpolation and the rendering quality. Let all the captured rays passing through A be denoted by Ω. In this case, Ω is a subset of all the known rays Θ. All of these rays in Ω potentially could be used for interpolation; however the rendering method has a *ray selection mechanism M* to choose a subset of rays ω from Ω to estimate the unknown ray r. Clearly: $\omega \subseteq \Omega \subseteq \Theta$.

Subsequently, an interpolation function F is applied to ω to estimate the value of the unknown ray r. Both M and F may or may not use some kind of scene geometric information G such as focusing depth or depth map. Mathematically, the LF rendering can be formulated as (1) and (2) below. Different LF rendering methods differ in their respective M and F functions and their auxiliary information G.

$$\omega = M(\Theta, G) \tag{1}$$
$$r = F(\omega, G) \tag{2}$$

Sampling Density (SD) is defined as the number of acquired rays per unit area of the scene space (number of rays in Ω divided by the area A) and Effective Sampling Density (ESD) as the number of rays per unit area of the scene that has been acquired and is employed during the interpolation process to estimate the unknown ray (number of rays in ω divided by the area A), that is,

$$SD = \frac{|\Omega|}{A} \tag{3}$$
$$ESD = \frac{|\omega|}{A} = \frac{|M(\Theta, G)|}{A} \tag{4}$$

where $|\Omega|$ and $|\omega|$ are the number of elements in Ω and ω respectively. A is the area of interpolation convex hull and can be calculated by deriving the line equations for the boundary rays β_i's and finding the vertexes of convex hull A at depth d.

It has been shown in [11, 12] that ESD is an indicator that can objectively determine the quality of an LF rendering method for a given LF acquisition configuration and scene. The higher the ESD, the higher the quality of the rendered video. Hence, for a target output quality, it is possible to determine the required ESD.

Sampling density SD is a parameter to quantify the acquisition. ESD is to quantify the combined effect of acquisition and rendering. Since $\omega \subseteq \Omega$ in any point of the scene space, ESD is less or at best equal to SD.

Fig. 1 demonstrates an LF rendering method with two-plane parameterization, camera plane uv and image plane st, using a depth map as the auxiliary information G. Ray r is the unknown ray that needs to be estimated for an arbitrary viewpoint reconstruction. r is assumed to intersect the scene on point p at depth d. If the exact depth d of point p is known; applying a back projection could easily find a subset of known rays Ω intersecting the scene at the vicinity of p. Subsequently, an adequate subset ω of these rays can be selected by mechanism M of the rendering method to be employed in interpolation F and r can be estimated as $r = F(\omega, G) = F(M(\Theta, G), G)$. If rays intersecting the scene at the vicinity of p don't pass through known pixel values in st, mechanism M will also select additional rays required for estimation of those rays with neighbourhood estimation or bilinear interpolation over st.

However, in practice, the estimated depth of p has an error Δd. This makes the rays intersect in an imaginary point p' in the space and going through the vicinity of area A on the scene instead of intersecting with the exact point p on the scene surface. Subsequently, this estimation error Δd would result in reduction of ESD and increase the distortion. To compute Ω in this case, back projection should be applied to the vertexes of A and not p to find all the rays passing through A.

The size of area A depends on Δd and as Δd gets larger, it also

increases. Usually only the upper bound of the error is known and therefore in this paper, the worst-case scenario, i.e., largest A is computed in the LF analysis which corresponds to the lower bound of ESD.

In Fig. 1, seven rays from all rays intersecting imaginary p are selected by M, i.e., $|\omega| = 7$, assuming these rays pass through known pixels or if neighbourhood estimation is used. In the case of bilinear interpolation in st plane, 28 rays are chosen by M to estimate these 7 rays. The chosen cameras in uv plane are bounded by a convex hull A'. It is easy to show that interpolation convex hull A is proportional to A'. Optical analysis of light field considers A' as the size of the light field synthetic aperture which defines the depth of view and focusing sharpness [10, 29].

Finally a 2D interpolation F over convex hull A' on uv plane can be applied to estimate unknown ray r from the rays in ω. This rendering method with depth information is referred to as UV-DM when 2D interpolation is performed over neighbouring cameras in the uv plane and neighbourhood estimation, i.e., choosing the closest pixel in the st plane. The rendering method is called UVST-DM in the case of 2D interpolation over neighbouring cameras in the uv plane and bilinear interpolation over neighbouring pixels in the st plane.

Notice that all the existing LF rendering methods such as [5-10], in which depth map is utilized, are a special case of UV-DM and UVST-DM methods. The ESD for the UV-DM and UVST-DM demonstrated in Fig, 1 can be derived as:

$$\text{ESD}_{\text{UVDM}} = \frac{|\omega|}{A} = \frac{|\omega|}{\frac{\Delta d}{d}A' + \mu(l(d+\Delta d), A')} \quad (5)$$

$$\text{ESD}_{\text{UVSTDM}} = \frac{|\omega|}{A} = \frac{|\omega|}{\frac{\Delta d}{d}A' + \mu(2l(d+\Delta d), A')} \quad (6)$$

where μ is a function to calculate the effect of pixel interpolation over st plane on the area A. A is mainly determined by A', but the pixel interpolation μ in (5) and (6) also has small effect on A. The pixel interpolation over st even when $\Delta d = 0$ makes $A = (ld)^2$. Note that to calculate A from A', the worst-case scenario is assumed, that is, the maximum value of Δd and the maximum area of A. This results in a lower bound for the ESD. Hence, the actual ESD varies from ideal ESD $= \frac{|\omega|}{(ld)^2}$ to the value calculated from (5) and (6).

In a simple form of UV-DM and UVST-DM, the rays in ω are selected in a way that A' becomes rectangular, i.e., 2D grid selection and therefore 2D interpolation over A' can be converted into a familiar bilinear interpolation. By further simplification for a regular camera grid and 2D grid selection of rays with 4 and 16 samples in $|\omega|$ respectively, (5) and (6) become:

$$\text{ESD}_{\text{UVDM}} = \frac{4}{\left(\frac{\Delta d.k}{d} + l(d+\Delta d)\right)^2} \quad (7)$$

$$\text{ESD}_{\text{UVSTDM}} = \frac{16}{\left(\frac{\Delta d.k}{d} + 2l(d+\Delta d)\right)^2} \quad (8)$$

where k is the distance between the two neighbouring cameras in the camera grid and l is the length of the pixel in the image plane as illustrated in Fig. 1. Note that most existing rendering methods with depth information adopt these simple versions of UV-DM and UVST-DM and choose only a very small subset of Ω, typically 4 or 16 rays, as ω. When the depth map is accurate, a small number of rays, say 4, would be sufficient, but for the case of less accurate depth maps, employing more rays in ω for interpolation could compensate for the adverse effect of errors in depth to some degree and improve the rendering quality since ESD is increased as can be seen from (5) and (6). This does not necessarily mean to increase the number of cameras, as there are

already $|\Omega|$ rays passing through area A of the scene and potentially can be chosen as ω. These samples are already captured so if using more can result in rendering quality improvement, the added complexity of the rendering algorithm may be justifiable.

For the rest of this paper, the analysis is only carried out for UV-DM, which can easily be extended to UVST-DM. Consider the simple form of UV-DM (i.e., the rays in ω are selected in a way that A' becomes rectangular). Mathematically, a general representation of this simplified UV-DM rendering method is $r = \text{UVDM}(d, \Delta d, k, l, |\omega|)$, where k is the distance between the two neighbouring cameras and l is the length of the pixel, d and Δd are the estimated depth and its error and $|\omega|$ refers to the number of rays selected by M and employed in interpolation F.

2.2. ESD for UVDM$(d, \Delta d, k, l, |\omega|)$

By extending (7), the ESD could be calculated for UVDM$(d, \Delta d, k, l, |\omega|)$ as follows:

$$\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)} = \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}(\sqrt{|\omega|}-1)\right)^2} \quad (9)$$

Equation (9) assumes that the rays are chosen for interpolation symmetrically around the vertical and horizontal axes, such as $4x4$ samples. In this case, $\sqrt{|\omega|}$ would be an integer. For an asymmetrical choice of rays, (9) could be rewritten as follow:

$$\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|_{vertical},|\omega|_{horizontal})} =$$
$$\frac{|\omega|_{vertical} \cdot |\omega|_{horizontal}}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}(|\omega|_{vertical}-1)\right)\left(l(d+\Delta d) + \frac{\Delta d.k}{d}(|\omega|_{horizontal}-1)\right)} \quad (10)$$

$\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)}$ predicts the rendering quality as shown in [11, 12]. In the above expression, d is given by scene geometry and Δd is determined by the depth estimation method and cannot be altered by us. Changing the other three parameters could potentially improve the rendering quality.

However, for a given acquisition configuration, k and l representing camera density and camera resolution are fixed and the only other parameter than can be tuned to compensate for error Δd while maintaining the rendering quality is $|\omega|$, the number of rays employed by the interpolation algorithm. Clearly, ESD is proportional to $|\omega|$, thus selecting more rays for interpolation results in a higher ESD value.

2.3. The Relationship between ESD and Number of Rays in $|\omega|$

Fig. 2 shows the theoretical calculation of mean $\overline{\text{ESD}}$ for UVDM$(d, \Delta d, k, l, |\omega|)$, for a given light field system with regular camera grid with $k = 5$ and $l = 0.01$, average depth of scene $\bar{d} = 100$, relative depth map error $\frac{\Delta d}{d}$ between 0% to 20%, for three different values of $|\omega| = 4, 16$ and 32. Notice that the estimation error for depth map in most real application is around 10% to 20%.

As can be observed from Fig. 2, higher errors in depth estimation result in less ESD and subsequently less rendering quality when $|\omega|$ is fixed. The reason is that error in depth Δd increases the area A for a given $|\omega|$ and therefore decreases ESD. However, choosing more rays for interpolation could increase the ESD and consequently rendering quality. For example, the ESD for 16-rays interpolation with errors less than 7% is still better than 4-rays interpolation with 1% error or ESD for 32-rays interpolation with errors less than 2% is still better than 16-rays interpolation with 1% error. However, for a very high level of errors in depth estimation, the ESDs in all three cases are declining rapidly to a

very small value and consequently the rendering quality may become inadequate.

This analysis shows that increasing the number of rays for interpolation could compensate for the adverse effect of depth map estimation errors on ESD to some degree, at least when the depth error is not very large. Of course, when more rays are employed in the interpolation, more computation is required. Thus in an LF rendering with a prior knowledge of the error bound in the depth map, the optimum number of rays $|\omega|$ could be calculated in advance.

2.4. Optimization of $|\omega|$

As discussed before, ESD is proportional to $|\omega|$. On the other hand, the complexity of interpolation is increased significantly with large $|\omega|$. Thus $|\omega|$ should be set at an optimum value to satisfy both the rendering quality and efficiency requirements. In this section, a theoretical minimum $|\omega|$ to compensate for the effect of errors in depth maps is derived. It is assumed that camera density is such that there is always enough number of rays in Ω to be used for interpolation.

In an ideal scenario, where there are no errors in depth map estimation and there is a depth map for each camera in the system, depending on the complexity of reflectivity of surfaces in the scene, one or more rays would be enough for an accurate rendering. In this case,

$$\text{ESD}_{\text{Ideal}} = \text{ESD}_{\text{UVDM}(d,0,k,l,n)} = \frac{n}{(ld)^2} \text{ and } n \geq 1 \quad (11)$$

where $n = 1$ is for the pure Lambertian reflection scene. Higher value of n can be used for non-Lambertian reflection.

So, the optimization problem is posed as follows: what would be the minimum $|\omega|$ (i.e., the minimum number of rays selected for interpolation by the *ray selection process M*) for any given UVDM($d, \Delta d, k, l, |\omega|$) with known depth map error Δd to have the same ESD as the ideal case?

$$\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)} = \text{ESD}_{\text{Ideal}} \rightarrow \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} =$$

$$\frac{n}{(ld)^2} \rightarrow |\omega| = \left(\frac{l(d+\Delta d) - \frac{\Delta d.k}{d}}{\frac{ld}{\sqrt{n}} - \frac{\Delta d.k}{d}}\right)^2 \quad (12)$$

where $k < \frac{ld^2}{\Delta d \sqrt{n}}$

Equation (12) gives the minimum $|\omega|$ required for interpolation in rendering process to avoid quality deterioration due to errors in depth maps.

For the purpose of this paper, it is assumed that available $|\Omega|$ and thus SD is always large enough to provide this minimum $|\omega|$ in each point of the scene.

It should be noted that ESD is a function of d, the depth of a point in the scene space. Hence, it has different values at different points of the scene. Therefore, typically for a given scenario, (12) is applied to the mean $\overline{\text{ESD}}$ for the entire scene by assuming the average depth of the scene $\bar{d}$ and average error in depth $\overline{\Delta d}$ to calculate average $\overline{|\omega|}$. Employing $\overline{|\omega|}$ rays in interpolation, guarantees the scene to be sampled and rendered with average $\overline{\text{ESD}_{\text{Ideal}}}$.

If the design criteria requires the scene to be sampled and reconstructed by a minimum $\text{ESD}_{\text{Ideal}}$ instead of average $\overline{\text{ESD}_{\text{Ideal}}}$, (12) should be applied to all d ranging between (d_{min}, d_{max}), the minimum and maximum depths of the scene with corresponding Δd. This gives optimum $|\omega|$ for each depth d and the maximum $|\omega|$ can be chosen by *ray selection mechanism M* of a rendering method.

Fig. 3 shows the same system demonstrated in Fig. 2, but this time for any $\Delta d < 20\%$, $|\omega|$ is calculated directly from (12) to maintain $\overline{\text{ESD}}$ at 4.00, the ideal ESD calculated for $n = 4$. k is calculated as follows to satisfy the condition of (12): $k < \frac{0.01 x 100^2}{20\sqrt{4}} < 2.5 \rightarrow k = 2.2$. Fig. 4 shows the actual number of rays $|\omega|$, employed in interpolation in such a scenario. The corresponding point for 10% error in depth estimation is highlighted in Fig. 3 and Fig. 4, respectively, to show the relation of these two Figures.

Note that ESD cannot be increased indefinitely by only increasing $|\omega|$ because: a) The curve of ESD vs. $|\omega|$ is saturating as $|\omega|$ increases: $\lim_{|\omega|\to\infty}(\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)}) = \left(\frac{d}{\Delta d k}\right)^2$, b) $|\omega|$ is bounded by $|\Omega|$ and cannot be increased indefinitely, i.e., ESD cannot be increased more than SD on any point of the scene because both $|\Omega|$ and thus SD are predetermined by the acquisition configuration, and c) Increasing $|\omega|$ would also increase the complexity of interpolation process significantly. Hence, in practice, the error in depth map can be compensated for by judicious alteration of *both* $|\omega|$ and k, i.e., higher rendering complexity and camera density.

3. EXPERIMENTAL VALIDATION

The main issue in quantitative analysis of LF rendering methods is the lack of ground truth data. To address this, a simulation system proposed in [30-32] was utilized. The simulator takes a 3D model of a scene and generates both reference cameras images and ground truth images. It also provides the depth maps for the following experiments. Controlled amount of depth map error is introduced to study how the rendering would be impacted when the depth map is noisy or inaccurate.

Fig. 5 illustrates the UV-DM rendering quality for four depth map error levels $\frac{\Delta d}{d} = 5\%, 10\%, 15\%, \text{and } 20\%$, and for each error level, different $|\omega| = 4, 9, 16, 25, \text{and } 36$. Thus, 20 different combinations of UVDM $(d, \Delta d, k, l, |\omega|)$ are demonstrated. Rendering quality is reported in terms of PSNR. Four different 3D scenes were chosen and a regular camera grid of $20x20$ was simulated as the LF acquisition component. For each experiment, 1000 random virtual cameras were produced. Each reported PSNR is averaged among 80,000 experiments for 1000 virtual cameras and four all 3D scenes.

As can be seen in Fig. 5, the same pattern expected from the proposed model is achieved, i.e., increasing the number of rays in interpolation improves the PSNR, e.g. $|\omega| = 25$ and 15% error performs better than $|\omega| = 9$ and 10% error.

3.1. Rendering with Desired PSNR

Assume the desired rendering quality is given as an average PSNR value. This section shows how the proposed optimization model can be used to calculate $|\omega|$ to produce the rendering quality at the desired PSNR value.

To be able to directly predict rendering PSNR from the theoretical ESD, an empirical relationship between calculated ESD and rendering PSNR values has been established:

$$\text{PSNR}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)} \cong$$

$$20 \log_{10} \frac{255}{\sqrt{Q.\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)}}^P} \quad (13)$$

where $1 < Q < 15$ and $-0.9 < P < -0.2$

Equation (13) is employed to calculate the corresponding ESD for a given PSNR value. Q and P for a given scene were approximated through experiments. Then (12) is applied to find the optimum number of rays $|\omega|$ to maintain the ESD and the corresponding PSNR at a prescribed value (for instance 50 dB), as shown in Fig. 6. Fig. 6 also shows the average PSNR for conventional fixed 4 rays interpolation, calculated number of rays $|\omega|$ is demonstrated in Fig. 7.

Fig. 6 shows that for high error rate, the use of optimum $|\omega|$ using (12) results in significant improvements over the conventional fixed 4 rays interpolation and can maintain the rendering quality around prescribed 50 dB. Fig. 6 and Fig. 7 are the experimental results corresponding to the theoretical predictions presented in Fig. 3 and Fig. 4.

The proposed method is also applied to real scenes for subjective evaluation. Two real scenes "Eucalyptus Flowers" and "Lego Knights" are chosen from Stanford light field archive [33]. A random 10% error in depth is applied to the depth maps. The proposed optimization method calculates optimum 12 and 14 rays interpolation for these scenes respectively to achieve desired rendering PSNR of 50 dB. Fig. 8 illustrates a sample rendering output for both scenes for 4, 8, and optimum 12/14 rays interpolation. As can be seen from Fig. 8, lower number of rays for interpolation results in blurry rendering. In contrast, employing optimum number of rays for interpolation results in all in focused and sharper rendering output. The reason is that, interpolation with lower number of rays is corresponded to smaller synthetic aperture size which results in higher depth of view but not sharp rendering for any point p in the scene. On the other hand, interpolation with higher number of rays is corresponded to larger synthetic aperture size which results in smaller depth of view but better focusing and sharper rendering for any point p in the scene. Note that as the depth of each p is known from depth map, the depth of view is not an important indicator for UV-DM rendering.

4. DISCUSSION AND CONCLUSION

An analytical model for evaluating the impact of depth map errors on rendering quality for LF based FVV systems based on ESD is presented in this paper. A method is developed from the model to calculate the optimum number of rays required for interpolation to compensate for the adverse effect of depth map errors on the rendering quality. To employ the proposed method in LF based FVV system design, the desired rendering quality of the system in PSNR can be mapped to the corresponding ESD by employing the empirical model given as (13). This ESD with depth estimation error is applied to (12) to calculate the optimum number of rays required for interpolation in rendering process.

5. REFERENCES

[1] E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*, pp. 3-20, 1991.

[2] M. Levoy and P. Hanrahan, "Light field rendering," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 31-42, Aug. 1996.

[3] S. J. Gortler, *et al.*, "The lumigraph," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 43-54, Aug. 1996.

[4] E. Camahort, *et al.*, "Uniformly sampled light fields," *Rendering Techniques*, vol. 98, pp. 117-130, 1998.

[5] K. Takahashi and T. Naemura, "Layered light-field rendering with focus measurement," *Signal Processing: Image Communication*, vol. 21, pp. 519-530, 2006.

[6] N. W. Daniel, *et al.*, "Surface light fields for 3D photography," in *Conference on Computer Graphics and Interactive Techniques*, 2000.

[7] Y. Jingyi, *et al.*, "Scam light field rendering," in *10th Pacific Conference on Computer Graphics and Applications*, 2002, pp. 137-144.

[8] H. Y. Shum, *et al.*, "Pop-up light field: An interactive image-based modeling and rendering system," *ACM Trans. Graphics*, vol. 23, pp. 143-162, Apr. 2004.

[9] W. Wen, *et al.*, "An efficient method for all-in-focused light field rendering," in *IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, 2010, pp. 399-404.

[10] I. Aaron, *et al.*, "Dynamically reparameterized light fields," in *Conference on Computer Graphics and Interactive Techniques*, 2000.

[11] H. Shidanshidi, *et al.*, "Objective evaluation of light field rendering methods using effective sampling density," in *MMSP*, 2011, pp. 1-6.

[12] H. Shidanshidi, *et al.*, "A Method for Calculating the Minimum Number of Cameras in a Light Field Based Free Viewpoint Video System," in *ICME*, 2013 pp. 1-6.

[13] J.X. Chai, *et al.*, "Plenoptic sampling," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 307-318, Jul. 2000.

[14] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 1038-1050, 2003.

[15] C. Zhang and T. Chen, "Light field sampling," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, pp. 1-102, 2006.

[16] L. Zhouchen and S. Heung-Yeung, "A Geometric Analysis of Light Field Rendering," *Int. J. Comput. Vision*, vol. 58, pp. 121-138, 2004.

[17] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, pp. 430-444, 2006.

[18] A. Pons, *et al.*, "Image quality metric based on multidimensional contrast perception models," *Displays*, vol. 20, pp. 93-110, 1999.

[19] S. Winkler, "A perceptual distortion metric for digital color images," in *ICIP*, 1998, pp. 399-403 vol. 3.

[20] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, pp. 2959-2965, 1995.

[21] T. Brandão and P. Queluz, "Towards objective metrics for blind assessment of images quality," in *ICIP*, 2006, pp. 2933-2936.

[22] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *ICASSP*, 2007, pp. I-869-I-872.

[23] S. Winkler, "Video quality and beyond," in *European Signal Processing Conference*, 2007, pp. 3-7.

[24] C. L. Zitnick, *et al.*, "High-quality video view interpolation using a layered representation," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 600-609, Aug 2004.

[25] S. M. Seitz, *et al.*, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *CVPR*, 2006, pp. 519-528.

[26] J. Kilner, *et al.*, "Objective quality assessment in free-viewpoint video production," *Image Commun.*, vol. 24, pp. 3-16, 2009.

[27] E. Bosc, *et al.*, "Towards a New Quality Metric for 3-D Synthesized View Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, pp. 1332-1343, 2011.

[28] E. Bosc, *et al.*, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," in *ICIP*, 2011, pp. 2597-2600.

[29] T. Feng and H. Y. Shum, "An optical analysis of light field rendering," in *Fifth Asian Conference on Computer Vision*, 2000, pp. 394-399.

[30] H. Shidanshidi, *et al.*, "A quantitative approach for comparison and evaluation of light field rendering techniques," in *ICME*, 2011, pp. 1-4.

[31] F. Safaei, *et al.*, "Scene-adaptive configuration of two cameras using the correspondence field function," in *ICME*, 2013, pp. 1-6.

[32] H. Shidanshidi, *et al.*, "Non-uniform sampling of plenoptic signal based on the scene complexity variations for a free viewpoint video system," in *ICIP*, 2013, pp. 3147-3151.

[33] S. U. Computer Graphics Laboratory. *The (New) Stanford Light Field Archive*. Available: http://lightfield.stanford.edu/lfs.html
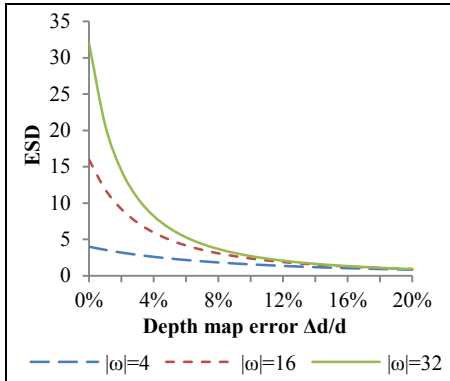
Fig. 2. Theoretical $\overline{ESD}$ for $UVDM(d, \Delta d, k, l, |\omega|)$ for $\bar{d} = 100$, depth map with relative error $\frac{\Delta d}{d}$ in the range of $[0\%, 20\%]$, for $|\omega| = 4, 16$ and 32
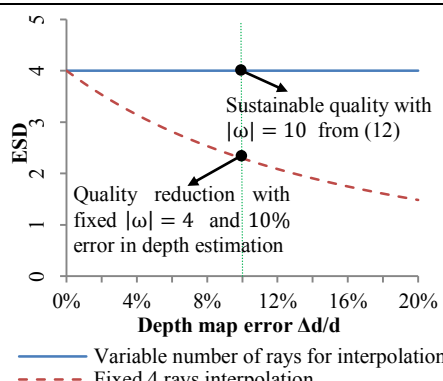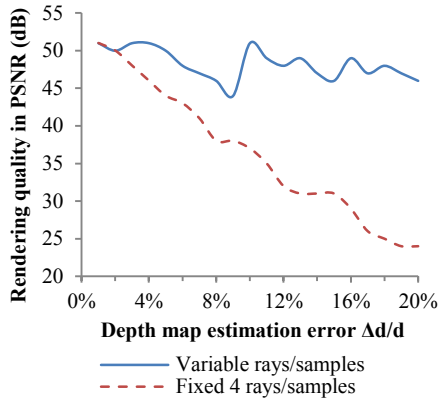

Fig. 3. Theoretical impact of depth estimation error on rendering quality ($\overline{ESD}$) for fixed $|\omega| = 4$ and calculated $|\omega|$ from (12)
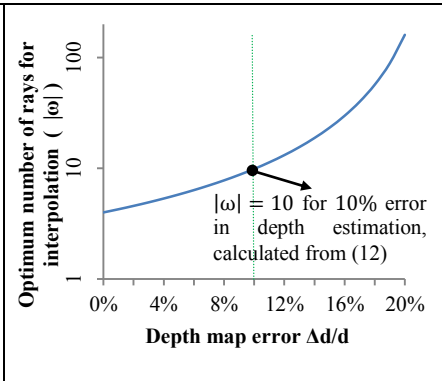
**Fig. 4.** Theoretical calculation of $|\omega|$ form (12) for different levels of errors to maintain the rendering quality ($\overline{ESD}$) at a constant value of 4.00
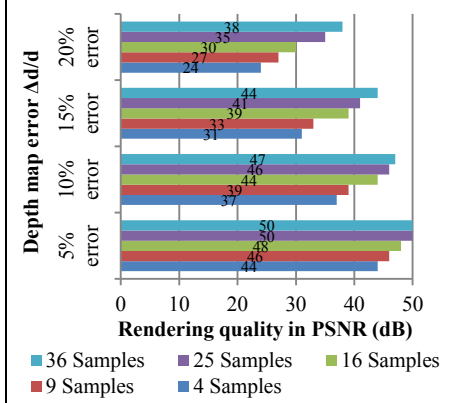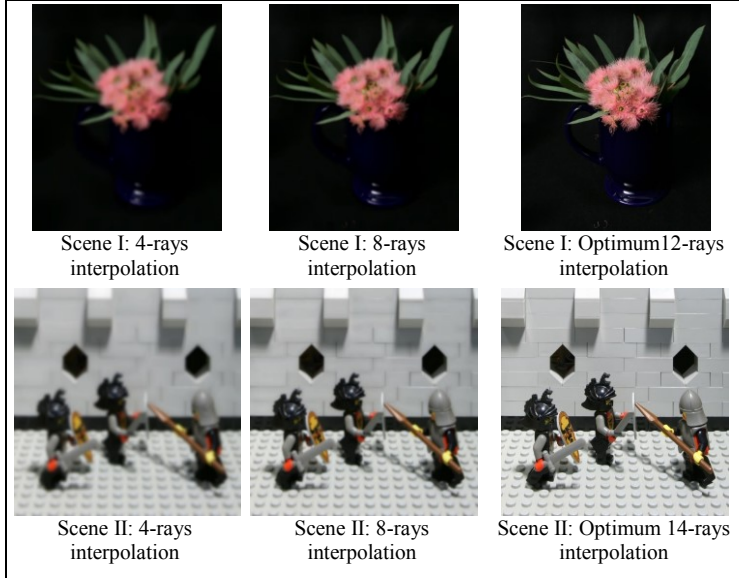

Fig. 5. Experimental $UVDM(d, \Delta d, k, l, |\omega|)$ rendering quality in PSNR for $\frac{\Delta d}{d} = (5\%, 10\%, 15\%,$ and $20\%)$ and $|\omega| = (4, 9, 16, 25,$ and $36)$


Fig. 6. Experimental rendering quality for conventional fixed 4 rays interpolation ($|\omega| = 4$) vs. calculated optimum number of rays $|\omega|$ demonstrated in Fig. 7 for different levels of depth errors
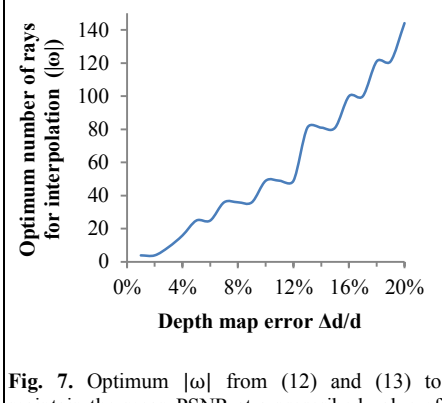
Fig. 7. Optimum $|\omega|$ from (12) and (13) to maintain the mean PSNR at a prescribed value of 50 dB for different levels of errors in depth estimation

Scene I: 4-rays interpolation

Scene I: 8-rays interpolation

Scene I: Optimum12-rays interpolation

Scene II: 4-rays interpolation

Scene II: 8-rays interpolation

Scene II: Optimum 14-rays interpolation

Fig. 8. Subjective evaluation of the proposed optimization method: Rendering output for two real scenes from Stanford data set [33], "Eucalyptus Flowers" and "Lego Knights" for 4 rays, 8 rays, and calculated optimum number of rays in interpolation (12 and 14 rays for scene I and II respectively)
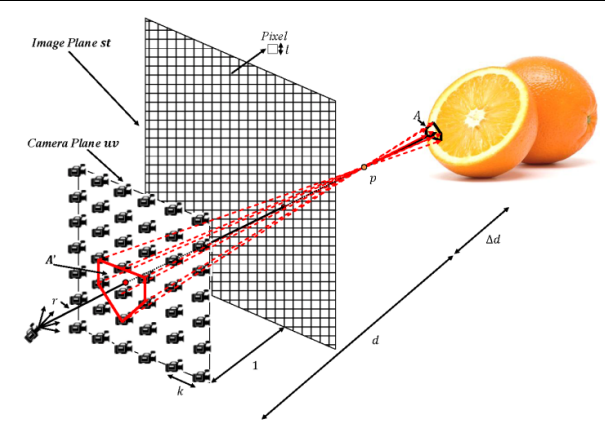
Fig. 1. General light field rendering method using depth information (UV-DM /UVST-DM) demonstrating unknown ray r, imaginary intersection point p with approximated depth d with Δd error in depth estimation, area of interpolation A, synthetic aperture A', pixel length l, distance between adjacent cameras k, image plane st, and camera plane uv.