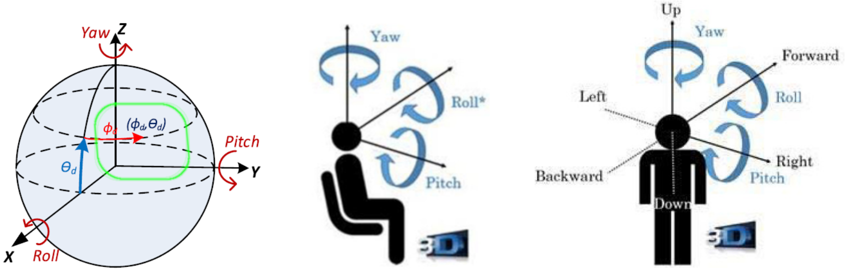
Displacement Error Analysis of 6-DoF Virtual Reality

Ridvan Aksu, Jacob Chakareski
The University of Alabama, Tuscaloosa, AL

Vladan Velisavljevic
University of Bedfordshire, Luton, UK



(a) User viewport and 6DoF.
(b) 3DoF (360°) video.
(c) 6DoF video.

Figure 1: Illustration of user viewport, degrees of freedom, and immersive video types [4].

ABSTRACT

Virtual view synthesis is a critical step in enabling Six-Degrees of Freedom (DoF) immersion experiences in Virtual Reality (VR). It comprises synthesis of virtual viewpoints for a user navigating the immersion environment, based on a small subset of captured viewpoints featuring texture and depth maps. We investigate the extreme values of the displacement error in view synthesis caused by depth map quantization, for a given 6DoF VR video dataset, particularly based on the camera settings, scene properties, and the depth map quantization error. We establish a linear relationship between the displacement error and the quantization error, scaled by the sine of the angle between the location of the object and the virtual view in the 3D scene, formed at the reference camera location. In the majority of cases the horizontal and vertical displacement errors induced at a pixel location of a reconstructed 360° viewpoint comprising the immersion environment are respectively proportional to 3/5 and 1/5 of the respective quantization error. Also, the distance between the reference view and the synthesized view severely increases the displacement error. Following these observations: displacement error values can be predicted for given pixel coordinates and quantization error, and this can serve as a first step towards modeling the relationship between the encoding rate of reference views and the quality of synthesized views.

KEYWORDS

Omnidirectional video, 6DoF, View synthesis, Depth-image-based rendering, Virtual Reality.

ACM Reference Format:
Ridvan Aksu, Jacob Chakareski and Vladan Velisavljevic. 2019. Displacement Error Analysis of 6-DoF Virtual Reality. In *13th International Conference on Distributed Smart Cameras (ICDSC 2019), September 9–11, 2019, Trento, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3349801.3349812

1 INTRODUCTION

The increasing popularity of immersive media and the latest developments in their effective capture, compression, streaming, and rendering push the development in remote immersion technologies. Following the recent popularity of 360° video and VR, more advanced immersive media formats are becoming the focus of the academia and the industry, e.g., Six Degrees of Freedom (DoF) video, Point Clouds, and Light Fields.

6DoF video, being the next generation video technology, uses multiple omnidirectional cameras to record a remote scene and provides freedom of movement in 3 rotational and 3 directional axes (Figure 1). Compared to 360° videos, which provides users a 360° look-around of a remote scene from a fixed location, 6DoF video additionally provides directional movement. This allows for both motion parallax, i.e., differentiation of far and close object's displacement by lateral head movement, and complete freedom within the spatial range of the scene. However, this freedom entails a heavy cost on the required data volume for storing, bandwidth for streaming, and computational intensity for rendering.

To achieve an immersive 6DoF experience, it is required to provide a continuous view corresponding to the movement of the user. That is, as a user moves in the scene, the view observed by the user should reflect that movement in a continuous manner. Since it is impossible to record a scene for every possible continuous viewpoint, for positions that do not coincide with camera locations, the selected view should be synthesized artificially from the available camera feeds. This can be achieved by using Depth-Image-Based Rendering (DIBR), which uses the views captured by existing cameras (*reference views*) and synthesizes the view at requested location (*the virtual view*). For DIBR process, both the texture of the objects and scene captured by the reference camera (*texture map*) and the distance of them (*depth map*) should be provided along with the positions of the virtual view and the reference view in 3D world. Currently the Reference View Synthesizer (RVS) software is used for view synthesis from multiple omnidirectional cameras, and is being standardized by MPEG for this purpose [13].

To freely observe a remote scene, texture and depth maps of multiple reference views should be available for user to perform

ICDSC 2019, September 9–11, 2019, Trento, Italy

ACM ISBN 978-1-4503-7189-6/19/09…$15.00
https://doi.org/10.1145/3349801.3349812

virtual view synthesis. Higher synthesized view quality requires more reference views with high quality depth and texture maps, resulting huge data rate requirements for 6DoF video.

From the original reference view at the server to the rendered virtual view at the client, signal distortion is introduced in several points. *Occlusion distortion*, *geometric distortion*, *encoding distortion of texture map*, and *displacement error* are four main distortion types in 6DoF video. The former two are due to the shape and positioning of the objects in the view and can be fixed by improving view synthesis algorithms for DIBR. The latter two are introduced during encoding and can play a critical role in video compression and quality trade offs. Encoded texture maps have similar rate-distortion characteristics of conventional videos. Encoding of depth maps, on the other hand, shifts the object positions in synthesized views and introduces a displacement error on object pixels. This shift indicates a unique rate-distortion dependency for 6DoF video.

In this study, we investigate the relationship between the depth map quantization error of the reference view and the displacement error in the synthesized view. Due to the quantization of the depth map pixels during encoding, depth intensity values are altered. This alteration, during the synthesis process, result in displacement of scene pixels in the synthesized image. This displacement alters the object shapes and texture, causing distortion in synthesized image. Understanding the displacement error is the first critical step in 6DoF rate-distortion modeling, the others being the displacement - video quality relationship and encoding - quantization modeling.

Multiple factors affect the displacement error during view synthesis: the distance x between the virtual view and the reference view in 3D, the distance d of the objects to the camera (i.e., depth intensity values), the position of the pixels in the depth map (ϕ, θ) with respect to the view synthesis direction $\hat{x}$, i.e, the vector from the anchor view position to the virtual view position, and the quantization error Δ. We analyze the generic view synthesis process to find an analytic dependency between these 6 factors and the displacement error of the synthesized view pixels.

The rest of the paper is structured as follows: In Section 2 we discuss the literature on 6DoF video and virtual view synthesis used in other media and the fundamentals of view synthesis. View synthesis of omnidirectional video is discussed in Section 3. Section 4 presents the displacement error modeling. We conduct experiments to verify our model and demonstrate our findings in Section 5. Finally, Section 6 concludes the paper.

2 BACKGROUND

In this paper we analyze the displacement error of 6DoF video based on various factors of generic view synthesis process. In the following, we first briefly discuss the structure of a 6DoF and the research status (Section 2.1). We then present a short background of virtual view synthesis and the pipeline of omnidirectional virtual view synthesis, which we use throughout the paper, in Section 2.2. Finally in Section 2.3 we introduce the projection performed between 3D and 2D representations of omnidirectional video.

2.1 6 DoF Video

In 6DoF video, a user observes a viewport of the recorded remote scene determined by Field of View (FoV) that follows user head orientation, at a point in that remote scene, which dynamically changes with her positioning. Figure 1 shows the axes and positioning used in immersive media and compares the regular omnidirectional video (3DoF) with 6DoF video. The user viewport (green window in Figure 1(a)) is determined by the position along (X,Y,Z) axes (directional positions) in the 3D scene, and centered by the Yaw, Pitch, and Roll angles (rotational positions around (X,Y,Z) axes). The Yaw and Pitch, ϕ_d and θ_d, corresponds to azimuth and polar angles of the spherical coordiantes. In 3DoF, only Yaw, Pitch, and Roll movements are possible, which alters the viewing direction of the user on a fixed point (Figure 1(b)). In 6DoF video, directional movements are also possible along (X,Y,Z) axes, which allows flexibility not only in the viewport direction but also in the viewport position (Figure 1(c)). In between these two concepts, 3DoF+ video is studied as an intermediate step, where a user can perform very limited directional movement and can achieve motion parallax.

Representation of a 6 DoF video faces several challenges. First, having multiple captured representations of a scene in high resolution requires enormous size of data. Then, it is possible to record the scene only from a limited small number of user view positions. For user positions with no recorded view, intermediate views can be virtually synthesized using existing recorded views. DIBR is a common way of synthesizing virtual views by using the depth map and texture map of the reference views, mainly used in multi-view images and videos [5]. Using the depth and texture maps of the reference view, first the 3D scene is constructed, then the corresponding texture map is synthesized from the 3D scene relative to the requested virtual view position. Also, it is usually required to have multiple reference views to synthesize an occlusion-free virtual view. So, while DIBR decreases the need of extensive recording of the scene with densely positioned cameras, requiring depth and texture maps of at least two reference views, increases the storage and delivery costs. An alternative to having explicit depth maps is using Structure from Motion (SfM) algorithms to derive 3D world coordinates of the scene, but this method is not suitable for all applications as it requires movement in the scene [9]. Another alternative is limited 6DoF video, where instead of continuous moving in the remote scene, user can be teleported between recorded camera positions [6]. Although this method is computationally less intensive and require less data volume, it offers a limited immersion.

The current focus regarding 6DoF video research is on improving view synthesis of 6DoF virtual view quality. For view synthesis, RVS [13] and VSRS360 [15] are the two state-of-the-art software specifically designed for 6DoF video. Kim et al. compared the quality of virtual views synthesized by RVS and VSRS360, showing the advantage of multiple reference view selection of RVS [11]. Jeong et al. performed a high level analysis of 3DoF video quality based on the distance and the quality of reference views and showed that symmetric down-sampling is better for equal distance and closer reference views while asymmetric down-sampling perform better if a reference view is further than the other [10]. Ray et al. investigated the view-synthesis based on projective cameras for 6DoF in order to exploit the parallax between divergent views and showed that it is possible to reach similar performances of omnidirectional cameras [14]. While these works provide insight about 6DoF video quality, they provide high-level analysis of reference view selection or camera positions. This is the first study on in-depth analysis of 6DoF rate-distortion modeling, specifically in displacement error.

2.2 Virtual View Synthesis

Virtual view synthesis uses an existing view of the scene from one position to reconstruct it in another position using the 3D geometry of the scene. For 3D video, multiple cameras on a straight line are used to record the scene, and the virtual views between cameras are synthesized to enable the continuous lateral user movement [2, 7, 17]. The distortion modeling of virtual views are investigated in [5] to find the optimal bit allocation for multi-view images. [3] proposed a scalable multi-view video streaming framework for multiple users, by employing multi-view distortion modeling, allowing low-cost dynamic multi-view streaming. However, due to the complexity of extra dimensions, this modeling is not applicable in 6DoF video.

Omnidirectional view synthesis is built on top of view synthesis of projective video. VSRS360 is based on VSRS [17], and RVS is built on SVS [8]. The main architecture of the both systems is as follows: first, the spherical coordinates of the 2D representation are calculated. Then, the depth maps recorded at the reference views are used to synthesize the depth and texture maps at the virtual view positions. Multiple virtual texture and depth maps originated from different reference views are blended, whereas the holes left because of occlusion are inpainted in a post-processing step.

2.3 3D World Projection

Omnidirectional view synthesis process requires projecting the 2D panorama (i.e., the 2D projected representation of the omnidirectional view) coordinates into 3D world coordinates. The equirectangular projection (ERP) is the most popular omnidirectional media format standardized by MPEG-I OMAF [1] and is acquired by projecting parallels and meridians of the 3D view sphere as perpendicular lines to a 2D plane. RVS uses ERP for input and output images and in this study we use ERP as well. As a direct result of using ERP, pixel coordinates in 2D plane are the polar coordinates of encircling spherical 3D view. This indicates that the resulting displacement error will be dependent on polar geometry.

In an equirectangular image with a size $w \times h$, spherical coordinates of the pixels (u, v) are calculated as: $\phi = -\pi + 2\pi\frac{u}{w}$, and $\theta = -\pi/2 + \pi\frac{v}{h}$. Here $\hat{r}$ is the vector direction from the camera to the object and it can be represented using spherical angles (ϕ, θ) as:

$$\hat{r} = \begin{bmatrix} \cos\phi\cos\theta \\ \sin\phi\cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1}$$

The position of the object in the 3D world is calculated as $x_{\texttt{world}} = r \times \hat{r}$ where r is the distance of the object.

The reverse projection from the 3D world coordinates to spherical angles is performed via the following: $\phi = \operatorname{atan2}(\langle\hat{r}, \hat{y}\rangle, \langle\hat{r}, \hat{x}\rangle)$ for azimuth angle and $\theta = \arcsin\langle\hat{r}, \hat{z}\rangle$ for polar angle.

3 EQUIRECTANGULAR VIEW SYNTHESIS

To calculate the displacement error, view synthesis process of equirectangular omnidirectional views should be examined. Before moving forward, it is important to clarify the direction and orientation of an omnidirectional camera. To define the position of objects with respect to camera, a direction is selected as front direction and spherical angles in (ϕ, θ) are assigned such that in the front direction $\phi = 0°$. Due to the radial symmetry of the equirectangular panorama, it is possible to rotate the whole 2D panorama in azimuth angle, without loss of generality, to align views to have the same orientation. For this reason, in our discussions, we assume, without loss of generality, all cameras have the same rotation angle.

To find the relative coordinates of objects in the target equirectangular panorama, 3D world coordinates of these objects should be calculated relative to the virtual view position. Since the origin of the previous 3D space is the reference view position, 3D coordinates (x, y, z) of each object should be updated to $(x - x_1, y - y_1, z - z_1)$ where (x_1, y_1, z_1) are the coordinates of the target view with respect to the reference view.

Using the 3D coordinates of pixels, we would like to calculate the spherical coordinates of the synthesized view r_1, ϕ_1, and θ_1. Let $r_0 = \sqrt{x^2 + y^2 + z^2}$ be the distance of a point on an object from the reference view, $r_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}$ be the distance from the target view and $x_{\texttt{world}} = (x, y, z)$ be the 3D world coordinates of that point with respect to the reference view. We can write r_1 as a function of spherical coordinates of the reference view (ϕ_0, θ_0, r_0) using (1) as: $r_1 = ((\cos\phi_0\cos\theta_0 r_0 - x_1)^2 + (\sin\phi_0\cos\theta_0 r_0 - y_1)^2 + (\sin\theta_0 r_0 - z_1)^2)^{1/2} = ((x_1 + y_1 + z_1)^2 - 2(x_1\cos\phi_0\cos\theta_0 + y_1\sin\phi_0\cos\theta_0 + z_1\sin\theta_0) + r_0^2)^{1/2}$. Furthermore, θ_1 and ϕ_1, the spherical coordinates of the projected point can be written as a function of spherical coordinates of the reference view and the coordinates of the target view: $\phi_1 = \arctan(\sin\phi_0\cos\theta_0 r_0 - y_1, \cos\phi_0\cos\theta_0 r_0 - x_1)$ and $\theta_1 = \sin^{-1}((\sin\theta_0 r_0 - z_1)/r_1)$.

We can simplify the coordinate system for the case when cameras are at the same height ($z_1 = 0$) by altering the alignment of the cameras during 2D to 3D conversion, without loss of generality. To achieve that, we set the front direction of the reference camera ($X - axis$) to be the vector from the reference camera to the target view. In this case, relative coordinates of the target camera becomes $(x_1, 0, 0)$. Since this rotation can be achieved by only shifting the equirectangular panorama pixels horizontally, the projection and synthesis itself are not affected. After the simplification:

$$\begin{aligned} r_1 &= \sqrt{x_1^2 - 2x_1\cos\phi_0\cos\theta_0 r_0 + r_0^2} \\ \phi_1 &= \arctan(\sin\phi_0\cos\theta_0 r_0, \cos\phi_0\cos\theta_0 r_0 - x_1) \\ \theta_1 &= \sin^{-1}(\sin\theta_0 r_0, r_1) \end{aligned} \tag{2}$$

4 DISPLACEMENT ERROR MODELING

When the reference depth images are encoded to achieve lower data rates, the reference depth map intensity values (d_0) are affected by quantization error (Δ). This quantization error has the following impacts on the synthesized view, which is analyzed in the sequel: 1) change in the depth value (d), 2) change in the horizontal pixel position (Δ_u), and 3) change in the vertical pixel position (Δ_v).

The change in the depth intensity value alters the depth values of the target view. The change in the pixel coordinates causes pixels to be horizontally or vertically misplaced on the target equirectangular panorama. As a result of these changes, texture pixel positions are altered, which results in incorrect object shapes and textures. This, eventually can be connected to the final synthesized image quality.

Following the simplified model from (2), we would like to represent the depth value error in synthesized view Δ_d, and the changes in azimuth and polar angles, Δ_ϕ and Δ_θ, respectively, due to the quantization error of the reference view as a function of several parameters. In particular, we are considering the original depth map pixel intensities d_0, applied quantization error to these pixels

Δ, the polar and azimuth angles of the original pixels θ_0 and ϕ_0, the distance between the reference and virtual views x_1, and the view synthesis rotation angle α. Due to the simplification in (2), the azimuth angles of the rotated panorama become $\phi_0 - \alpha$ by aligning the camera to the view synthesis direction.

To achieve this model, we first derive the spherical coordinates ϕ_1', θ_1' and the depth value r_1' of the virtual view pixels synthesized from an encoded reference view. According to MPEG depth map format standards [16], a depth value r in normalized disparity format can be written as $d = \frac{1/r - 1/r_{\max}}{1/r_{\min} - 1/r_{\max}}.d_{\max}$. By taking $r_{\max}$ in infinite, it can be simplified to $d = d_{\max}.\frac{r_{\min}}{r}$ We start by the relationship of Δ_r and Δ on the relationship of depth d_0 and r_0 and then derive the rest of the dependencies:

$$\Delta = d_0' - d_0, \quad r_0' = 255\frac{r_{\min}}{d_0 + \Delta}, \quad r_0 = 255\frac{r_{\min}}{d_0} \tag{3}$$

The depth values of the target view pixels after quantization error can be written as following by using (3) in (2):

$$r_1' = \sqrt{x_1^2 - 2(\frac{255 r_{\min}}{d_0 + \Delta})(x_1 \cos\phi_0 \cos\theta_0) + (\frac{255 r_{\min}}{d_0 + \Delta})^2} \tag{4}$$

Finally, the resulting quantization error in the circular coordinates of the target system can be written as: $\Delta_\phi = \text{arctan2}(\sin\phi_0 \cos\theta_0 r_0, \cos\phi_0 \cos\theta_0 r_0 - x_1) - \text{arctan2}(\sin\phi_0 \cos\theta_0 (r_0 + \Delta_r), \cos\phi_0 \cos\theta_0 (r_0 + \Delta_r) - x_1)$ and $\Delta_\theta = \sin^{-1}(\sin\theta_0 r_0 / r_1) - \sin^{-1}(\sin\theta_0 (r_0 + \Delta_r)/r_1')$. Using arctan($a$)–arctan($b$) and arcsin($a$)–arcsin($b$) relationships we can reach te following equations, where r_1 and r_1' are defined in (2) and (4) respectively:

$$\begin{aligned}
\Delta_\phi =& \arctan(\sin\phi_0 \cos\theta_0 \frac{255 r_{\min}}{d_0 + \Delta} x_1, x_1^2 \\
& - \cos\phi_0 x_1 (\frac{255 r_{\min}(2d_0 + 3\Delta)}{d_0(d_0 + \Delta)}) + \cos^2\theta_0 \frac{255 r_{\min}}{d_0}(\frac{255 r_{\min}}{d_0 + \Delta})) \\
\Delta_\theta =& \sin^{-1}((\sin\theta_0 (\frac{255 r_{\mathtt{min}}}{d_0} \sqrt{(r_1')^2 - \sin^2\theta_0 (\frac{255 r_{\min}}{d_0 + \Delta})^2} \\
& - \frac{255 r_{\min}}{d_0 + \Delta} \sqrt{(r_1^2) - \sin^2\theta_0 (\frac{255 r_{\mathtt{min}}}{d_0})^2}) / r_1' r_1)
\end{aligned} \tag{5}$$

In addition to the angular displacement, we can calculate the pixel displacement in the 2D panorama by substituting Δ_ϕ and Δ_θ with $\Delta_u = \frac{w}{360}\Delta_\phi$ and $\Delta_v = \frac{h}{180}\Delta_\theta$. Finally, we can write Δ_d, the depth error of virtual view, as $\Delta_d = 255 \cdot r_{\min}/(r_1' - r_1)$.

5 EXPERIMENTATION

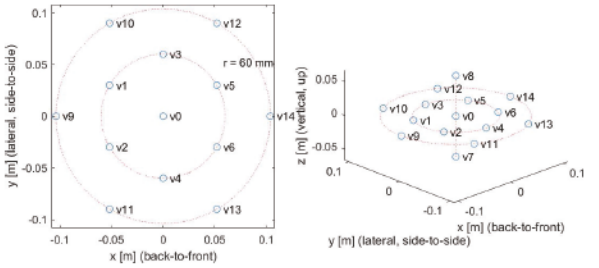


Figure 2: Positions of cameras in Classroom sequence

We investigated the parameters of ClassroomVideo [12] dataset, in order to determine the valid intervals of the reference view and view synthesis parameters. Specifically, we investigated the upper and lower bound, average, and standard deviation values of quantization error Δ, depth value d_0, distance of the reference and synthesized views x_1, view synthesis rotation angle α, and pixel position angles θ_0, ϕ_0 in the ClassroomVideo dataset. Table 1 shows the upper and lower bounds of the reference view and view synthesis parameters with their average values (where applicable).

The dataset includes a total of 15 4K 360° cameras with depth and texture maps. Cameras are positioned on two concentric circles. One camera placed at the center of the circles, and two more are placed above and below it (Figure 2).

To encode the reference depth maps, two compression modes are used: finer and coarser. In the finer mode, the encoding Quantization Parameter (QP) was set to 22, whereas, in the coarser mode, it was set to 42. In both modes, the magnitude of the depth map quantization value ($|\Delta|$) has the same average value of 1.8. Even in coarser mode, the standard deviation is small, indicating that the magnitude of the quantization error is less than 5. This effectively limits the range of $|\Delta|$ for the majority of pixels. The average depth value of 106 and the standard deviation of 46, implies that all depth values between 0-255 are being used. In this dataset, the distance between cameras x_1 is between 0.06 m and 0.18 m. The influence of the distance x_1 on the synthesized view quality has been previously analyzed in part in [10] showing that the smaller values of x_1 benefit the quality. However, this relation has not been quantified so far up to our best knowledge.

Parameter	Min	Max	Avg	Std
ϕ_0	−180°	180°	-	-
θ_0	−90°	90°	-	-
Width	1	4096	-	-
Height	1	2048	-	-
d_0	0	255	106	46
$\|\Delta\|$ for QP=22	0	13	1.8	0.55
$\|\Delta\|$ for QP=42	0	70	1.8	1.88
x_1	0.06	0.18	0.12	-
α	−180°	180°	0	-

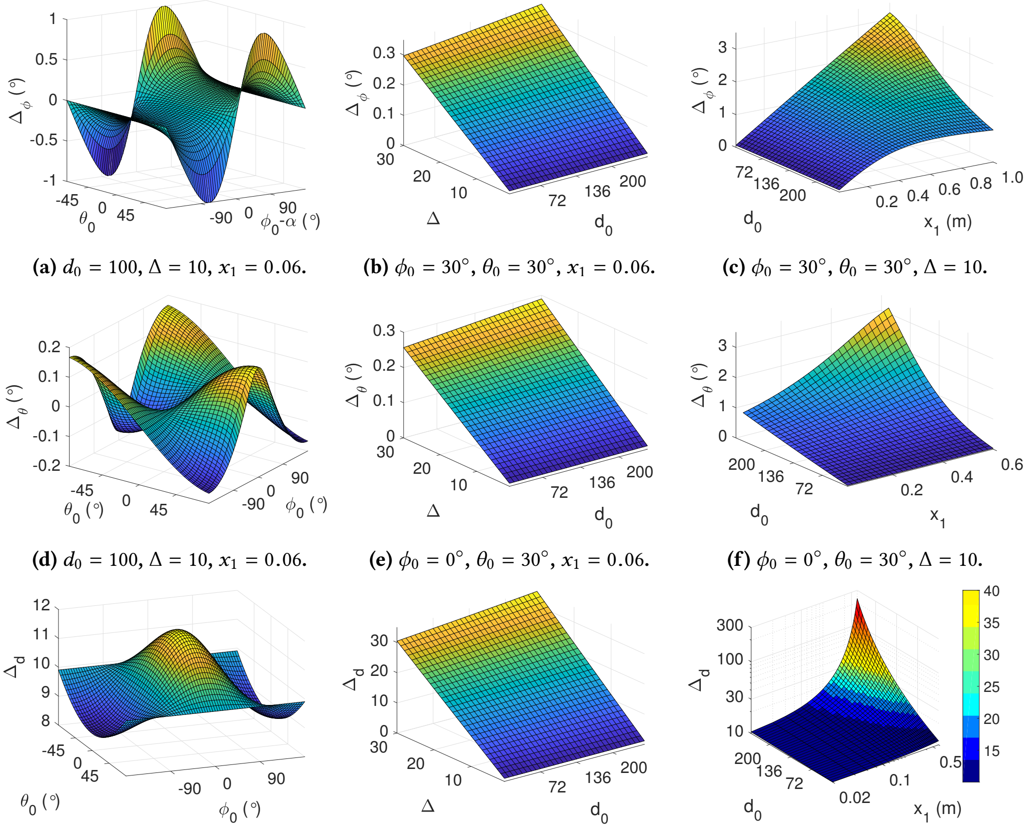
Table 1: Upper and lower parameter bounds in ClassroomVideo.

5.1 Displacement on Horizontal Axis

We start by examining the displacement error observed in azimuth angle of the synthesized depth view. The displacement error Δ_ϕ is the difference between azimuth angles of the synthesized views of the encoded and original depth maps. We analyze the bounds of Δ_ϕ and Δ_u in various regions of the 2D panorama, with varying reference view and view synthesis parameters.

In order to analyze the effect of the 6 parameters (Δ, d_0, ϕ_0, θ_0, x_1, and α) on Δ_ϕ, we set four of the parameters as constant and vary the others. Here α represents the rotation of the cameras and $\phi_0 - \alpha$ indicates the horizontal pixel coordinates of the reference depth map with respect to view synthesis direction.

First, we analyze the effect of the initial pixel coordinates on the quantization error. For that, we introduce a constant quantization error of $\Delta = 10$ at constant depth value $d_0 = 100$, to pixels residing in intervals $\phi_0 = (-180, 180)$ and $\theta_0 = (-90, 90)$. Figure 3a shows the resulting horizontal displacement Δ_ϕ in terms of degrees. In this figure, for illustration purposes, the bounds of θ_0 are set as $(-80°, 80°)$, since for $|\theta_0| > 80°$ the value of Δ_ϕ increases exponentially and becomes intractable. Three main observations can be



(a) $d_0 = 100, \Delta = 10, x_1 = 0.06$.

(b) $\phi_0 = 30°, \theta_0 = 30°, x_1 = 0.06$.

(c) $\phi_0 = 30°, \theta_0 = 30°, \Delta = 10$.

(d) $d_0 = 100, \Delta = 10, x_1 = 0.06$.

(e) $\phi_0 = 0°, \theta_0 = 30°, x_1 = 0.06$.

(f) $\phi_0 = 0°, \theta_0 = 30°, \Delta = 10$.

(g) $d_0 = 100, \Delta = 10, x_1 = 0.06$.

(h) $\phi_0 = 0°, \theta_0 = 30°, x_1 = 0.06$.

(i) $\phi_0 = 0°, \theta_0 = 30°, \Delta = 10$.

Figure 3: Analysis of Δ_ϕ, Δ_θ, and Δ_d by various parameters.

made from this figure: 1) The displacement error on the horizontal axis is minimum for the pixels that are in the view synthesis direction ($\phi_0 - \alpha = 0$) and maximum around the orthogonal directions to the view synthesis ($\phi_0 - \alpha = \pm 90$). This indicates that view synthesis direction and selection of the reference cameras have a significant effect on synthesis quality. 2) The displacement error towards the poles is higher compared to the error around the equator, where the error is minimized. The polar areas of the panorama have less significance on image quality since a large stretch represents a relatively small area on the original 3D sphere, thus a large displacement over the poles is likely to have a limited effect on synthesized image quality. 3) Displacement of pixels ranges between (-2,2) around the equator and increases to the range of (-6,6) around $\theta_0 = 70°$ (75th percentile of the video). This results in a small horizontal displacement over the majority of the video.

Second, we extend this to investigate the effect of changing d_0 and Δ on Δ_ϕ. Figure 3(b) shows this for $\theta_0 = 30°$ and $\phi_0 - \alpha = 90$. Here we observe that the effect of Δ can be approximated linearly, increasing with the Δ value. This is mostly due to the However, the effect of d on Δ_ϕ, while being linear, is negligible. This results in a simplified approximation of expected error based on Δ.

Finally, we analyze the effect of x_1 on the horizontal displacement in Figure 3(c). We see that the linear relationship between d_0 and Δ_ϕ becomes non-linear for increasing x_1. Note that, for smaller d_0, (i.e., distant objects) there is a linear increase in Δ_ϕ value, implying a steady increase in displacement error with the increase of distance between views. We observe that the horizontal displacement error increases up to 3° or a total of 40 pixels for a panorama with 4096 pixel width. For closer objects, this error converges to 1° or 11 pixels. This shows that while closer objects are subjected to a limited displacement error, the error on further objects increase with distance between cameras.

In summary, the impact of the quantization error is limited on the horizontal displacement in non-polar regions of the equirectangular panorama. In polar regions, where the ERP distortion peaks, error also becomes intractable. This suggests that a projection method with more contained distortion might respond better to quantization. For non-polar regions ($\theta_0 = (-70°, 70°)$) a quantization error of 10 results in a maximum 0.5° or 6 pixel shift suggesting a very limited horizontal displacement in synthesized view, considering a 4K resolution. Finally, the effect of quantization error on the horizontal displacement linearly increases with x_1, affecting the synthesized view quality severely at large x_1 values.

5.2 Displacement on Vertical Axis

Figure 3(d) shows the effect of pixel coordinate on the vertical displacement. Note that, compared to the horizontal displacement, the vertical displacement is more limited. Due to sinusoidal shape of both ϕ_0 and θ_0, the peak displacement error is observed in the polar areas, while error around the equator is limited. Even at the poles, the maximum error is very small: in pixel values, the maximum shift is about 2 pixels for $\Delta = 10$. As a result, the vertical pixel displacement is bounded by the 1/5th of Δ for a 4K video.

The effect of d_0 and Δ on Δ_ϕ has a linear trend similar to the horizontal case as seen in Figure 3(e). Here, the error is still bounded by $\pm 0.2^\circ$ or ± 2 pixels at $\Delta = 10$, for all d_0 values and pixel positions.

In the vertical displacement case, the effect of camera distance is more noticeable (Figure 3(f)). Especially when d_0 value is small (i.e., for closer objects) the vertical displacement increases rapidly with the increase of camera distance. The error becomes significant when camera distance is above 0.2 m. This is occurring because the vertical pixel shift gradually increases to 1° (11 pixels) from 0.2° (2 pixels), and then rapidly after that point.

In summary, the maximum vertical displacement is observed on poles with about 1/5 pixel shift to error ratio. It decreases with the cosine of θ to the equator and becomes intractable with large distance between cameras.

5.3 Depth Value Error

Unlike the displacement error, the depth error has an indirect effect on synthesized view quality. A large depth error on synthesized view implies possible ghosting artifacts by altering the depth values of foreground and background and by putting background objects in front of foreground objects.

Figure 3(g) illustrates the effect of initial pixel coordinates. While the Δ_d is around Δ ($\Delta = 10$), it scales with $\sin\phi_0 \sin\theta_0$. So, the maximum error is observed on the equator and in the view synthesis direction, while the minimum error is observed on the equator but in the opposite direction of view synthesis.

In terms of quantization error value and initial depth value, depth error linearly follows the quantization error as seen in above two parameters (Figure 3(h)). Figure 3(i) illustrates the effect of the camera distance. Here we observe the most substantial effect of x_1 compared to the other cases. While around 0.2 meters, it is easy to contain the error (up to $\Delta_d = 20$ at $\Delta = 10$), it exponentially increases with the distance and becomes above 256 (8-bit depth value limit) around 0.5 meters. As a result, it is difficult to measure the depth error with cameras far apart than 0.2 meters and impossible around 0.5 meters.

6 CONCLUSION

We investigated the displacement error in 6DoF video based on the quantization error of the encoded reference view depth maps. Specifically, we analyzed the effect of the reference view parameters (quantization error, horizontal and vertical pixel coordinates, and depth value) and view synthesis parameters (virtual and reference view distance and orientation in 3D scene) on the synthesized view displacement error. This is the first step of modeling the reference view rate and virtual view quality in 6DoF video

We observed that both the horizontal and vertical displacements are linear with quantization error. Both displacement errors are bounded by quantization error and scales as sinusoidal functions of azimuth and polar angles, especially in non-polar regions. For the majority (75%) of the panorama there pixel shift to quantization error intensity is less than 3/5 in the horizontal direction and less than 1/5 in the vertical direction for a 4K 6DoF video. Direction of view synthesis plays an important role in all displacement errors: the minimum horizontal and the maximum vertical error is observed around the virtual view synthesis direction, that is the vector from the reference camera to the target view. The vertical depth error is also a linear function of the quantization error while its intensity varies with $\sin\phi_0 \sin\theta_0$ of the reference view pixel location. Finally, the distance of the reference and target views have a limited effect on all 3 variables for small distances. However, above 0.2 meter distance between cameras, the error values become intractable.

We found that the regions more prone to displacement error (e.g., the polar regions, regions perpendicular to view synthesis direction), the regions subject to higher quantization (e.g., object boundaries), and using far reference views require a finer Quantization Parameter (QP) value for encoding. The next step is investigating the effect of displacement error on the quality of the synthesized view texture map to provide a complete model of 6DoF video reference view encoding rate and the synthesized view quality.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (NSF) under Award CCF-1528030, Award ECCS-1711592, Award CNS-1836909, and Award CNS-1821875, and by research gifts and an Adobe Data Science Award from Adobe Systems.

REFERENCES

[1] 2017. Information Technology-Coded Representation of Immersive media (MPEG-I) - Part 2: Omnidirectional Media Format. document N17563, ISO/IEC JTC1/SC29/WG11 MPEG.

[2] B. Ceulemans, S. Lu, G. Lafruit, and A. Munteanu. 2018. Robust Multiview Synthesis for Wide-Baseline Camera Arrays. *IEEE Transactions on Multimedia* 20, 9 (Sep. 2018), 2235–2248.

[3] J. Chakareski, V. Velisavljevic, and V. Stankovic. 2013. User-Action-Driven View and Rate Scalable Multiview Video Coding. *IEEE Transactions on Image Processing* 22, 9 (Sep. 2013), 3473–3484.

[4] M. L. Champel, R: Koenen, G. Lafruit, and M. Budagavi. 2018. Draft 1.0 of TR: Technical Report on Architectures for Immersive Media. document N17685, ISO/IEC JTC1/SC29/WG11 MPEG, 123rd Meeting.

[5] G. Cheung, V. Velisavljevic, and A. Ortega. 2011. On Dependent Bit Allocation for Multiview Image Coding With Depth-Image-Based Rendering. *IEEE Transactions on Image Processing* 20, 11 (Nov 2011), 3179–3194.

[6] X. Corbillon, F. De Simone, G. Simon, and P. Frossard. 2018. Dynamic Adaptive Streaming for Multi-viewpoint Omnidirectional Videos. In *Proc. 9th ACM Multimedia Systems Conference (MMSys)*. Amsterdam, Netherlands.

[7] M. Domanski, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, J. Stankowski, R. Ratajczak, and T. Grajek. 2013. High Efficiency 3D Video Coding Using New Tools Based on View Synthesis. *IEEE Transactions on Image Processing* (Sep. 2013).

[8] S. Fachada, D. Bonatto, A. Schenkel, and G. Lafruit. 2018. Depth iamge based view synthesis with multiple reference views for virtual reality. In *2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. Helsinki, Finland.

[9] J. Huang, Z. Chen, D. Ceylan, and H. Jin. 2017. 6-DOF VR videos with a single 360-camera. In *2017 IEEE Virtual Reality (VR)*. Los Angeles, CA.

[10] J.B. Jeong, D. Jang, J. Son, and E.S. Ryu. 2018. 3DoF+ 360 Video Location-Based Asymmetric Down-Sampling for View Synthesis to Immersive VR Video Streaming. *MDPI Sensors* 18, 9 (2018).

[11] H.Y. Kim, Y.J. Lee, and J.G. Kim. 2018. Performance Analysis on View Synthesis of 360 Video for Omnidirectional 6DoF. In *Korean Broadcasting Engineering Society*.

[12] B. Kroon. 2018. 3DoF+ test sequence ClassroomVideo. document M42415, ISO/IEC JTC1/SC29/WG11 MPEG.

[13] B. Kroon and G. Lafruit. 2018. Reference View Synthesizer (RVS) Manual 3.0. document N18068, ISO/IEC JTC1/SC29/WG11 MPEG.

[14] B. Ray, J. Jung, and M. Larabi. 2018. On the possibility to achieve 6-DoF for 360 video using divergent multi- view content. In *2018 26th European Signal Processing Conference (EUSIPCO)*.

[15] K. Wegner, D. Losiewicz, T. Grajek, O. Stankiewicz, A. Dziembowski, and M. Domanski. 2018. Omnidirectional View Synthesis and Test Images. In *2018 International Conference on Signals and Electronic Systems (ICSES)*. Krakow, Poland.

[16] K. Wegner, O. Stankiewicz, T. Grajek, and M. Domanski. 2017. Depth map formats used within MPEG 3D technologies. document N16730, ISO/IEC JTC1/SC29/WG11 MPEG.

[17] K. Wegner, O. Stankiewicz, M. Tanimoto, and M. Domanski. 2013. Enhanced view synthesis reference software (VSRS) for free-viewpoint television. document M31518, ISO/IEC JTC1/SC29/WG11 MPEG.