View Synthesis: LiDAR Camera versus Depth Estimation

Yupeng XIE, Sarah FACHADA, Daniele BONATTO, Mehrdad TERATANI, Gauthier LAFRUIT

Université Libre de Bruxelles
LISA department
Av. F.D. Roosevelt 50 CP165/57
Belgium, 1050 Brussels

Yupeng.Xie@ulb.be; Sarah.Fernandes.Pinto.Fachada@ulb.ac.be; Daniele.Bonatto@ulb.ac.be;
Mehrdad.Teratani@ulb.ac.be; Gauthier.Lafruit@ulb.ac.be

Abstract

Depth-Image-Based Rendering (DIBR) can synthesize a virtual view image from a set of multiview images and corresponding depth maps. However, this requires an accurate depth map estimation that incurs a high computational cost over several minutes per frame in DERS (MPEG-I's Depth Estimation Reference Software) even by using a high-class computer. LiDAR cameras can thus be an alternative solution to DERS in real-time DIBR applications. We compare the quality of a low-cost LiDAR camera, the Intel Realsense LiDAR L515 calibrated and configured adequately, with DERS using MPEG-I's Reference View Synthesizer (RVS). In IV-PSNR, the LiDAR camera reaches 32.2dB view synthesis quality with a 15cm camera baseline and 40.3dB with a 2cm baseline. Though DERS outperforms the LiDAR camera with 4.2dB, the latter provides a better quality-performance trade-off. However, visual inspection demonstrates that LiDAR's virtual views have even slightly higher quality than with DERS in most tested low-texture scene areas, except for object borders. Overall, we highly recommend using LiDAR cameras over advanced depth estimation methods (like DERS) in real-time DIBR applications. Nevertheless, this requires delicate calibration with multiple tools further exposed in the paper.

Keywords

View Synthesis, Depth Estimation, LiDAR, Camera Calibration, DERS, DIBR, RVS

1 INTRODUCTION

Depth-image-Based-Rendering (DIBR) technology [7] is widely used in end-to-end immersive autostereoscopy, promoting the continuous progress of 3D computer vision applications [13] [14]. It uses multiviews and their associated depth maps to synthesize a realistic virtual view. MPEG-I (The Moving Picture Expert Group Immersive) has specially introduced its DIBR-based Reference View Synthesizer (RVS) [11], which can support view synthesis in real-time with a large baseline. However, as DIBR-based, RVS performance is highly dependent on its input depth maps quality.

DERS [15] can estimate high accuracy depth maps but with a high computational cost due the complexity of its algorithm, which remains a significant challenge for the real-time application purpose. Additionally, low texture regions of the image significantly reduce the algorithm's performance.

LiDAR-based RGB-D cameras currently play a significant role in the research field of computer vision [17]. Their high accuracy of depth map acquiring with the low computational cost promotes the development of real-time 3D applications [5]. Hence, using an RGB-D camera, such as the Intel Realsense LiDAR L515 (hereafter LiDAR), can be considered an alternative solution to DERS in real-time view synthesis applications.

Nonetheless, LiDAR's depth maps are difficult to evaluate due to the depth sensor's limitations to capture non-reflective colors, absorbing materials, and objects with light deflecting shapes [9]. Moreover, camera calibration and depth registration are required because the captured depth map and its associated color image have different resolutions and misalignment due to different sensors' (RGB and Depth) positions.

This paper proposes a method to evaluate the LiDAR camera's performance. Instead of assessing its depth maps accuracy directly, we evaluate the quality of the virtual view synthesized by RVS with depth maps and their corresponding color images of the LiDAR, which is precisely calibrated. We compare the quality of virtual views, synthesized using the depth maps of LiDAR, and the estimated one by DERS, respectively. In this

DOI:10.24132/CSRN.2021.3101.35

ISBN 978-80-86943-34-3

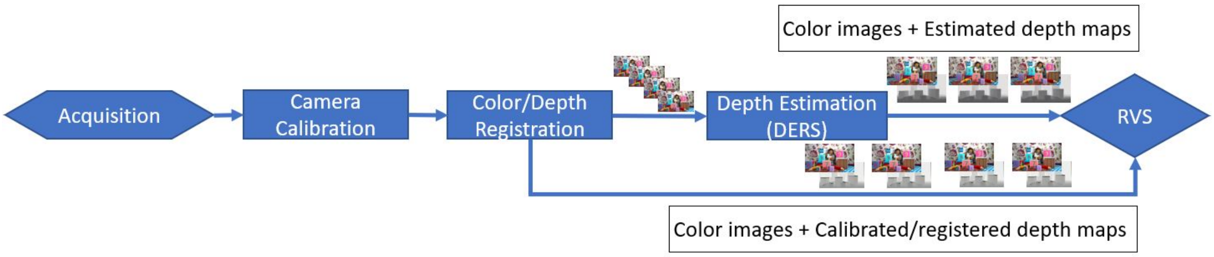
Color images + Estimated depth maps
Acquisition
Camera Calibration
Color/Depth Registration
Depth Estimation (DERS)
RVS
Color images + Calibrated/registered depth maps

Figure 1: Processing pipeline

comparison, we used the IV-PSNR virtual reality quality assessment metrics to evaluate their performance gap. Based on the evaluation results, we has observed the substantial trade-off between view synthesis performance using RVS, given the depth map acquired by LiDAR and estimated depth map by DERS. The processing pipeline is illustrated in Figure 1, which details are explained in section 2 and 3.

2 OVERVIEW OF THE PROCESSING PIPELINE

This section explains the processing pipeline illustrated in Figure 1. We also give a brief introduction to DERS, RVS, and our assessment measure IV-PSNR. The motivation and detailed procedure for camera calibration and depth registration are provided separately in section 3.

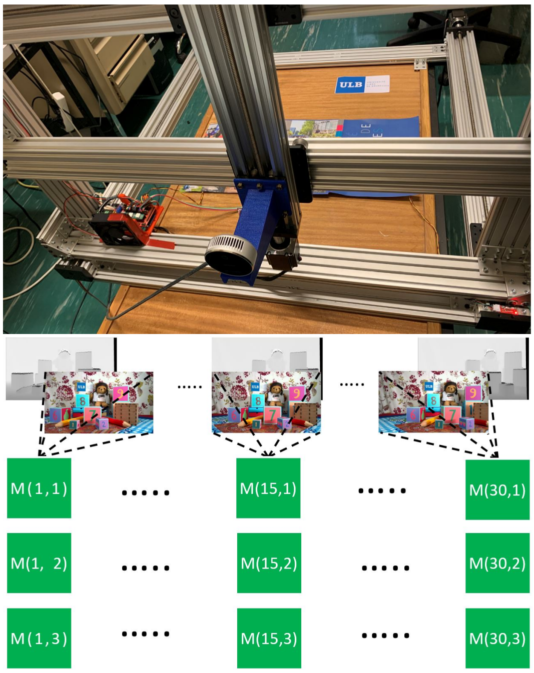

Figure 2: Acquisition system and 2D multiview configuration.

Acquisition

Before acquiring the test sequence for our evaluation of depth map quality by DERS and LiDAR for view synthesis, we have conducted a precise camera calibration, which is one of the contributions to this paper (section 3). Once the LiDAR was calibrated, we have mounted it on our acquisition robot to acquire a 2D dataset of 30x3 multiview color images and simultaneously registered their depth maps in real-time (Figure 2). The color images are used in DERS for estimation of depth maps and RVS for view synthesis.

Depth Estimation Reference Software (DERS)

DERS is one of the state of art high quality depth estimation software that have been promoted as reference software in MPEG-I. The latest version is 9.0 [12]. The main process of DERS is to estimate one depth map from a sparse setup of multiple reference color images. The algorithm includes two fundamentals steps.

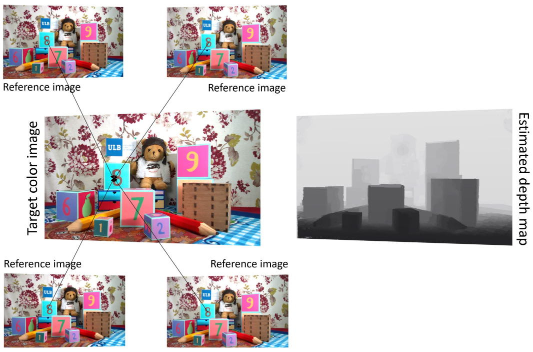

Figure 3: Estimation of depth map using five reference views.

(a) Calculate matching cost cube: DERS works in depth or disparity estimation mode. For the sake of simplicity, we consider the procedure to estimate disparities. The disparities can then be transformed into depth maps using [1]. Disparity estimation is performed using pair of images: The reference and the evaluation image. Both images are registered in a preprocessing step. Each pixel of the reference image corresponds to only an unknown pixel in the evaluation image. DERS performs a Sum of Absolute Differences (SAD) between a patch around the pixel in the reference image and all the patches centered on the corresponding epipolar line in the evaluation image to find matching candidates.

This procedure gives rise for each pixel of the reference image to a cost function on all the possible disparities along the epipolar line. Each cost function per pixel is then stored in a cost cube with dimensions width × height × cost. With the cost value varying between the minimum disparity and the maximum disparity. This procedure is performed between all images and the reference image, and the resulting cost volumes are merged and stored in a matching cost cube.

(b) Graph cut global optimisation: Selecting the best cost for each pixel in the cost cube results in noisy depth maps. Therefore, DERS performs a global optimization technique known as Graph-cut [3] [4] to obtain the cost cube's optimal values. Furthermore, a smoothing map is used to increase or decrease the cost of a cut between the image's pixels to consider if two pixels are on the same object. This increases the optimization quality by forcing close-by pixels to have similar depth values.

Figure 3 demonstrates one example of the DERS depth map, which is estimated by using five reference (including the target one in the center) color images. Subjectively, the depth map's quality is clean and sharp without outliers. However, it is not accurate to conclude before evaluating the quality of the virtual view synthesized by RVS using the estimated depth map.

Reference View Synthesizer (RVS)

RVS and it's real time extension RaVIS [2] is a DIBR-based software developed in the context of MPEG-I standardization activities. It has been designed to take any number of reference images, with corresponding depth maps, in any configuration and renders outputs by interpolation and extrapolation. Using several reference images makes it more resistant to DIBR artifacts such as ghosting (due to poor calibration) and disocclusions (due to missing information in the input images).

IV-PSNR

For evaluating quality metrics of RVS output, we have used IV-PSNR [6] which is a PSNR-based quality metric, which defined by equation (1) and (2). IV-PSNR takes YUV [16] (Y defines the luma component and two chrominance components U-blue projection and V-red projection) file as its input. On YUV images, the IV-PSNR is the weighted mean on each component:, where MAX is the maximum possible pixel value. Similarly to the MSE, $IVMSE$ is the mean of the squared error IVE for each pixel p of the virtual view, where D is a correction term taking into account the global color difference between the virtual and reference image.

$$IVPSNR_{yuv} = 10 \times log\left(\frac{MAX^2}{IVMSE}\right) \tag{1}$$

$$IVMSE = \sum_{y=0}^{H}\sum_{x=0}^{W} Min_{p_{R(x,y)}\in\Omega} \frac{(p_V(x,y) - p_R(x,y) + D)}{WH} \tag{2}$$

Unlike classic PNSR, IV-PSNR considers a patch of adjacent pixel quality metric evaluation instead of each single pixel. In this respect, it less considers the corresponding pixel shift in the objects' edges and is insensitive to the global color's difference between the ground truth and virtual view, which is suitable for immersive video quality metric evaluation. Please refer to the reference [6] for more details.

3 LIDAR CALIBRATION AND REGISTRATION

This section presents one of this paper's main contributions to explain how to use LiDAR correctly. Multiple camera calibration is requisite for adequately using the LiDAR camera to capture test sequence or view synthesis. The depth maps need to be aligned to their corresponding color image since they are captured separately from two sensors Figure 4.

The performance of such registration relies on the sensor's intrinsic and extrinsic parameters accuracy. However, the default parameters provided by the Intel SDK are too coarse to conform RVS's particular requirement. Hence, we have used Kalibr [8] [10] to calibrate the LiDAR camera for getting more robust parameters of sensors.

Calibration: Kalibr is a conventional calibration software that supports multiple camera calibration with a non-global field of view (do not restrict entire calibration target captured in each sensor). Its implementation is relatively straightforward and outputs a detailed calibration statement to help users understand the calibration accuracy.

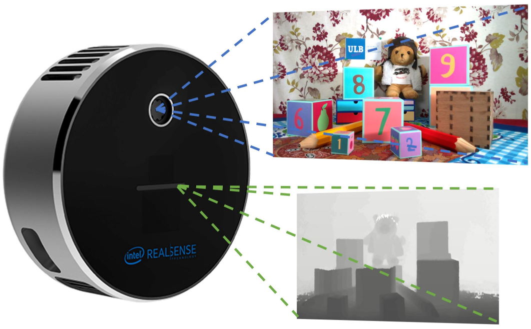

Figure 4: Intel L515 camera layout

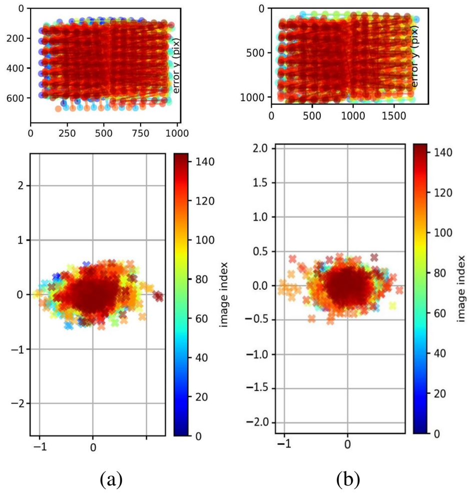

Figure 5: Calibration accuracy: (a) Distribution of detected pixel of the depth sensor and the interval of its related reprojection error.(b) Distribution of detected pixel of the color sensor and the interval of its related reprojection error.

We use a calibration pattern that is attached on a flat glass. This is important to achieve the high accuracy of calibration. According to the Kalibr's output Figure 5 (a and b), the LiDAR's depth sensor reprojection error is reduced to under ±0.5 pixel for the color sensor and ±1 pixel for the depth sensor. Compared with the uncalibrated registered depth maps (hereafter LiDAR-UCRD), the calibrated registered depth maps (hereafter LiDAR-CRD) have significantly impacted the view synthesis's quality (detailed discussion are in section 4).

Depth Map Registration: We have performed the depth registration process based on all calibration parameters. Nevertheless, it still has inevitable imperfections (Figure 6) even though camera calibration was sufficiently precise.

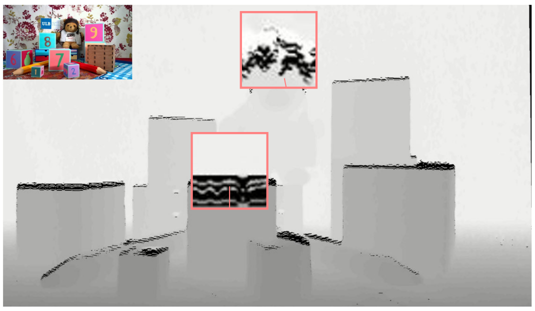

Figure 6: LiDAR registered depth map. (**a**) Error area above the cube. (**b**) Error area at the top of bear's head

In Figure 6 (**a**), the black pixels are the typical artifacts in registered depths. Depth sensor has a lower resolution than the color sensor, which cause some missing pixels in calibrated and registered LiDAR (LiDAR-CRD). Moreover, the color depth is located on top of the depth sensor. When projecting back depth map corresponding 3D points to the color sensor, some pixels information are missing since these specific pixels occluded in the depth sensor view.

In Figure 6 (**b**), the blank depth area in front of the bear's hat comes from the material's absorption of the light. LiDAR does not receive any reflected of a light ray in this area, which leads to invalid depth information.

Having LiDAR calibrated and registered, we can use its depth map to synthesize virtual views by RVS.

4 EXPERIMENTS

We have used two different RVS (Figure 7 and Table 1). Figure 7(a) demonstrates using the four reference images plus corresponding depth maps from 4 corners (i.e., $M_r(1,1)$, $M_r(30,1)$, $M_r(1,3)$, $M_r(30,3)$) of the dataset to synthesize 15 different intermediate virtual views (i.e., $M_v(1,2)$ to $M_v(15,2)$) by RVS with a large baseline-15cm. As illustrated in Figure 7(b),we used

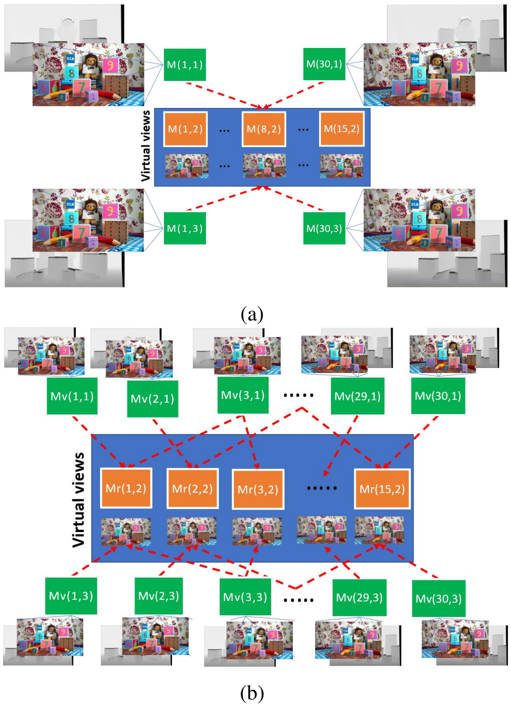

Figure 7: Experimental configuration using RVS

Table 1: Experimental configuration using RVS

RVS setup	DERS/LiDAR Camera
No. of input (image+depth)	4
Color image resolution	1920x1080
Depth resolution	1920x1080
Large baseline	15cm
Small baseline	2cm

RVS to synthesize the same virtual views but with a small baseline of 2cm with the inputs selected orderly from four adjacent reference images plus corresponding depth maps. This process accordingly was repeated with different depth maps DERS, LiDAR-UCRD, and LiDAR-CRD. In our evaluation, we have combined objective evaluation by IV-PSNR and subjective evaluation by examining the virtual views' quality.

Objective Evaluation

LiDAR-UCRD vs. LiDAR-CRD: Figure 8 (a) and (b) show the performance evaluation of the virtual view quality with LiDAR CRD and LiDAR UCRD in IV-PSNR, based on different baseline setup to RVS (Figure 7). The 15 virtual views ($M_{(1->15,2)}$) performance with LiDAR-CRD (blue line) is slightly higher than with LiDAR-UCRD (red line). Nevertheless, these gaps are nearly negligible since IV-PSNR less sensible to pixel shifting, and this can not show the quality improvement by precise camera calibration. Therefore, we have used subjective evaluation to approve the benefits of improving virtual view quality with LiDAR-CRD, demonstrated in the following subsection (Subjective Evaluation).

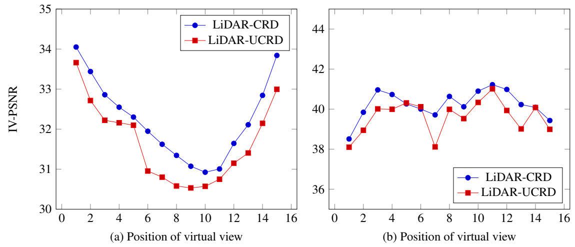

Figure 8: IV-PSNR (virtual view), LiDAR-UCRD vs. LiDAR-CRD. (a) Virtual views $M_{((1->15,2)}$ (15cm BL); (b) Virtual views $M_{(1->15,2)}$ (2cm BL)

LiDAR-CRD vs. DERS: According to Figure 9 (a) and (b) the virtual view synthesis using LiDAR CRD $M_{(1->15,2)}$ maintain at least 32.2dB with a large 15cm baseline or 40dB with a small 2cm baseline in IV-PSNR are acceptable values. Compared with the DERS-based (green line) virtual views, the latter outperforms about 4dB.

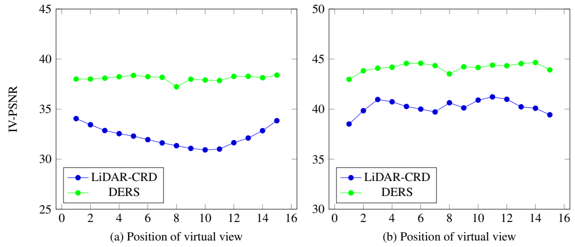

Figure 9: IV-PSNR (virtual view), LiDAR-CRD vs. DERS. (a): Virtual views $M_{(1->15,2)}$ (15cm BL); (b): Virtual views $M_{(1->15,2)}$ (2cm BL)

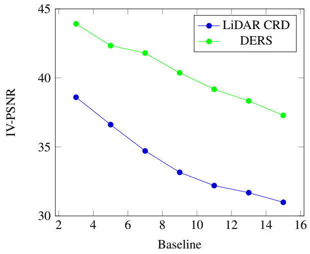

Figure 10: IV-PSNR (virtual view), LiDAR-CRD vs. DERS: For viewpoint M(8,2), when the baseline distance among reference views varies

Figure 10 shows the virtual views' with DERS and LiDAR-CRD given a viewpoint at M(8,2), while the baseline distance for the references images are varied. For one dedicated intermediate virtual view $M_{(8,2)}$, DERS-based virtual view has a constant superiority in IV-PSNR than LiDAR-based virtual view.

Subjective Evaluation

Referring to the objective measures presented above, we cannot fully conclude precisely. To have a better understanding of the advantages and disadvantages of LiDAR vs. DERS, in the following, we compare them subjectively and report their computational performances.

Figure 11(1th row) demonstrates one of ground-truth (original image) and outputs of RVS based on three different depth maps of DERS, LiDAR UCRD, LiDAR CRD. We have used the error map Figure 11(3rd row) to better demonstrate the differences between each virtual view and the ground truth. In the error map, brighter pixel color means a more significant error; and vice versa.

LiDAR-UCRD vs. LiDAR-CRD: Fig, 11(c), (d) and (4th row) show that the virtual view quality with LiDAR CRD significantly better than with LiDAR-UCRD. Thanks to the precise calibration and the accurate depth registration, the virtual view has fewer pixels shifting in the objects.

LiDAR-CRD vs. DERS: The error map of Figure 11(3rd row) shows the main reason for higher IV-PSNR in DERS compared to LiDAR-CRD. LiDAR camera suffers from the typical weakness in boundary scanning. The LiDAR CRD-based view virtual has some objects with border shrinking according to Figure 11((b), (d), in the first two columns of 4th row). Following three reasons can explain these differences: **1) The excessive incident Angle of the laser**: Too wide intersection angle between the incident and reflected

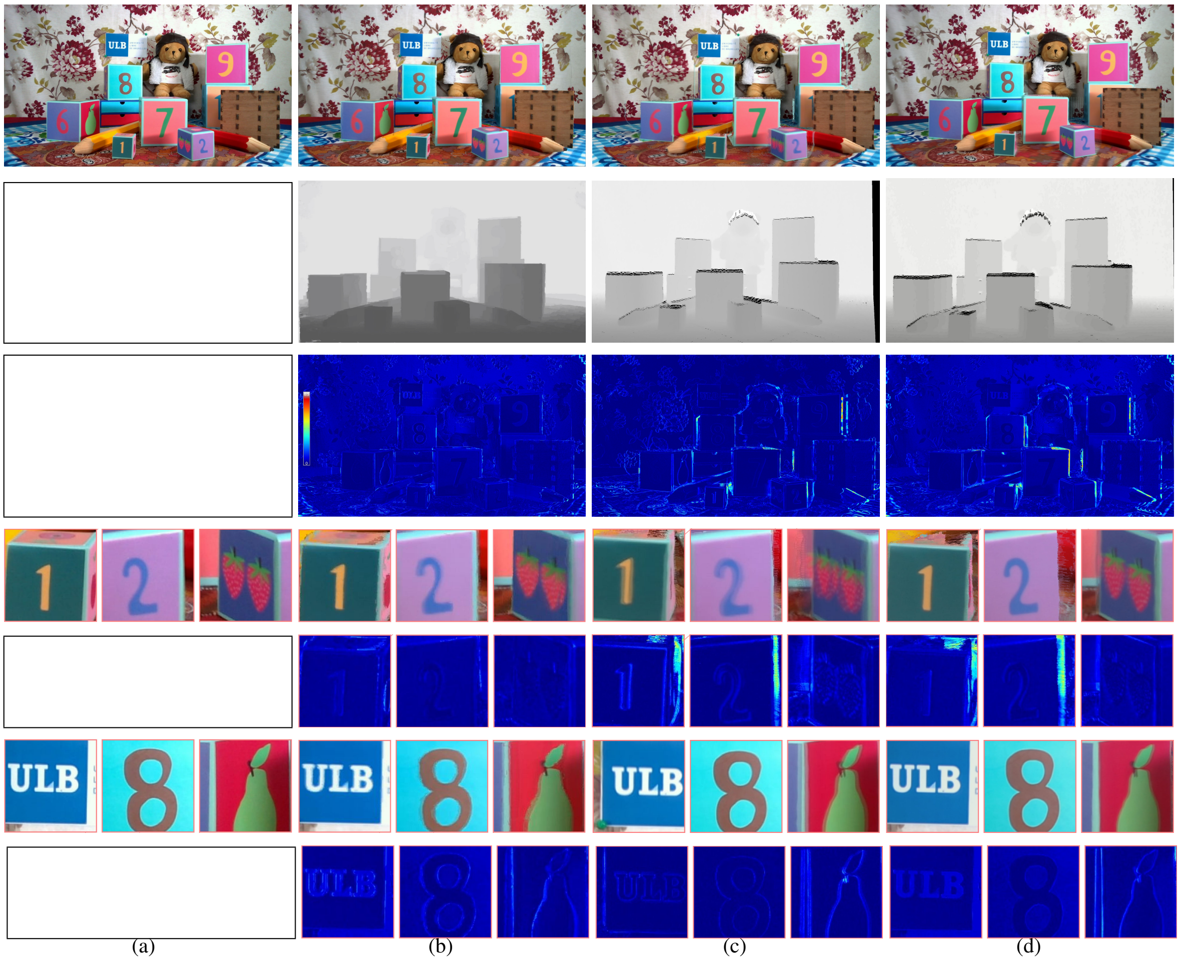

Figure 11: Subjective evaluation: (a) Ground-Truth, for column (b), (c) and (d), the first-row is DERS-based virtual view, second-row is LiDAR-UCRD-based virtual view, third-row is LiDAR-UCRD-based virtual view and the other rows are magnified regions and corresponding error maps.

ray that hampers LiDAR camera to get receiving echoes properly. **2) Registered depth map occlusion**: The occlusion problem of depth registration, which is vertically shifted in this LiDAR. **3) Laser interference**: The superfluousness of depth scanning among complex structural scenes, echoes interfere with each other at the high multi-reflection area. This interference brings some errors into depth information measurement.

However, DERS also has its vulnerability. Its performance is usually worse in low texture areas because its cost matching step relies on changing pixel value intensity. Low texture areas make cost matching difficult that prominently affects DERS output accuracy. Due to this flaw, the DERS-based view has some objects with pixels shifting. In contrast, LiDAR cameras do not suffer from getting wrong depth information in low texture areas as they benefit from its active acquisition attribute. The differences shown in Figure 11((b), (d), 5th row)

Last but no least, we compare the computation performance required for acquiring depth by LiDAR and estimating by DERS. DERS requires extremely high computational resources than the LiDAR according to the Table 2. The run time easily reaches around 4minutes to estimate one depth map with a high-class PC. Therefore, it is generally not possible to use DERS for a real-time DIBR system purpose.

Table 2: Comparison of the computation costs between DERS and LiDAR

PC Configuration	i9-10900X, 64GB Ram	
	DERS	LIDAR
CPU usage	90%	6% ∼ 30%
Ram usage	12GB	250Mb
Processing time	4.1 mins	Real time (30fps)

5 CONCLUSION

Using the accurately calibrated and precisely registered depth maps of the Intel Realsense LiDAR camera can output sufficiently good virtual views by RVS. When compared with the DERS-based virtual views, the latter outperforms the former objectively in IV-PSNR. We have subjectively observed that DERS only outperforms LiDAR in border areas. LiDAR camera's depth maps have better performance than DERS in low texture with no border area. Both DERS and the LiDAR camera's depth maps have a similar performance in high texture with no border area. Therefore, overall, the LiDAR camera's depth maps showed a substantial trade-off to DERS in virtual view quality subjectively. Moreover, DERS requires remarkably high computational resources, and its processing run time is relatively long, up to several minutes per depth map estimation. In contrast, using the LiDAR camera, without any computational cost, can achieve an acceptable trade-off in subjective quality in comparison with DERS. Therefore, we recommend the RGB-D camera such as LiDAR for real-time DIBR application purposes.

6 ACKNOWLEDGMENTS

This paper has been supported by the EU project HoviTron (Holographic Vision for Immersive Tele-Robotic OperatioN), Call identifier: H2020-ICT-2019-3, Grant Agreement: 951989. Sarah Fachada is a research fellow at FNRS (Fonds de la recherche scientifique). Authors appreciate the support of Prof. Eduardo Juarez, in cross checking the achieved results by Mr. Jaime Sancho, at UPM, Spain.

REFERENCES

[1] S.R. Barry and O. Sacks. *Fixing My Gaze: A Scientist's Journey Into Seeing in Three Dimensions*. Basic Books, 2009. ISBN: 9780786744749.

[2] Daniele Bonatto, Sarah Fachada, and Gauthier Lafruit. In: *RaViS: Real-time accelerated View Synthesizer for immersive video 6DoF VR*. Vol. 2020. Jan. 2020, pp. 382–1. DOI: 10.2352/ISSN.2470-1173.2020.13.ERVR-381.

[3] Y. Boykov, O. Veksler, and R. Zabih. In: *Fast approximate energy minimization via graph cuts*. Vol. 23. 11. 2001, pp. 1222–1239. DOI: 10.1109/34.969114.

[4] Y. Y. Boykov and M. -. Jolly. "Interactive graph cuts for optimal boundary region segmentation of objects in N-D images". In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 1. 2001, 105–112 vol.1. DOI: 10.1109/ICCV.2001.937505.

[5] Yan-Pei Cao, Leif Kobbelt, and Shi-Min Hu. "Real-time High-accuracy Three-Dimensional Reconstruction with Consumer RGB-D Cameras". In: *ACM Transactions on Graphics* 37 (Sept. 2018), pp. 1–16. DOI: 10.1145/3182157.

[6] Adrian Dziembowski. In: *Software manual of IVPSNR for Immersive Video*. ISO/IEC JTC 1/SC 29/WG 04 N0013, (Oct. 2020).

[7] Christoph Fehn. "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV". In: *Stereoscopic Displays and Virtual Reality Systems XI*. Ed. by Mark T. Bolas et al. Vol. 5291. International Society for Optics and Photonics. SPIE, 2004, pp. 93–104. DOI: 10.1117/12.524762.

[8] Paul Furgale, Joern Rehder, and Roland Siegwart. In: *Unified temporal and spatial calibration for multi-sensor systems*. Nov. 2013, pp. 1280–1286. DOI: 10.1109/IROS.2013.6696514.

[9] Ying He et al. "Depth Errors Analysis and Correction for Time-of-Flight (ToF) Cameras". In: *Sensors* 17 (Jan. 2017), p. 92. DOI: 10.3390/s17010092.

[10] Jörn Rehder et al. In: *Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes*. 2016, pp. 4304–4311.

[11] A. Schenkel S. Fachada D. Bonatto and G. Lafruit. "Depth image 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)*. 2018, pp. 1–4. DOI: 10.1109/3DTV.2018.8478484.

[12] S.Rogge et al. In: *MPEG-I Depth Estimation Reference Software*. 2019, pp. 1–6. DOI: 10.1109/IC3D48390.2019.8975995.

[13] M. Tanimoto et al. In: *Free-Viewpoint TV*. Vol. 28. 1. 2011, pp. 67–76. DOI: 10.1109/MSP.2010.939077.

[14] M. P. Tehrani et al. In: *Free-viewpoint image synthesis using superpixel segmentation*. Vol. 6. June 2017. DOI: 10.1017/ATSIP.2017.5.

[15] Krzysztof Wegner and Olgierd Stankiewicz. In: *DERS Software Manual*. ISO/IEC JTC1/SC29/WG11 M34302, (07. 2014).

[16] Ning Xu and Yeong-Taeg Kim. "Luminance preserving color conversion for 24-bit RGB displays". In: *2009 IEEE 13th International Symposium on Consumer Electronics*. 2009, pp. 271–275. DOI: 10.1109/ISCE.2009.5156864.

[17] Michael Zollhöfer et al. "State of the Art on 3D Reconstruction with RGB-D Cameras". In: *Computer Graphics Forum* 37 (May 2018), pp. 625–652. DOI: 10.1111/cgf.13386.