

# Depth Augmented Omnidirectional Stereo for 6-DoF VR Photography

Tobias Bertel<sup>\*†\*</sup> Moritz Mühlhausen<sup>\*‡§</sup> Moritz Kappel<sup>§</sup> Paul M. Bittner<sup>§</sup> Christian Richardt<sup>\*</sup> Marcus Magnor<sup>§</sup>  
<sup>§</sup> Institut für Computergraphik, TU Braunschweig, Germany  
<sup>\*</sup> Visual Computing Group, University of Bath, UK

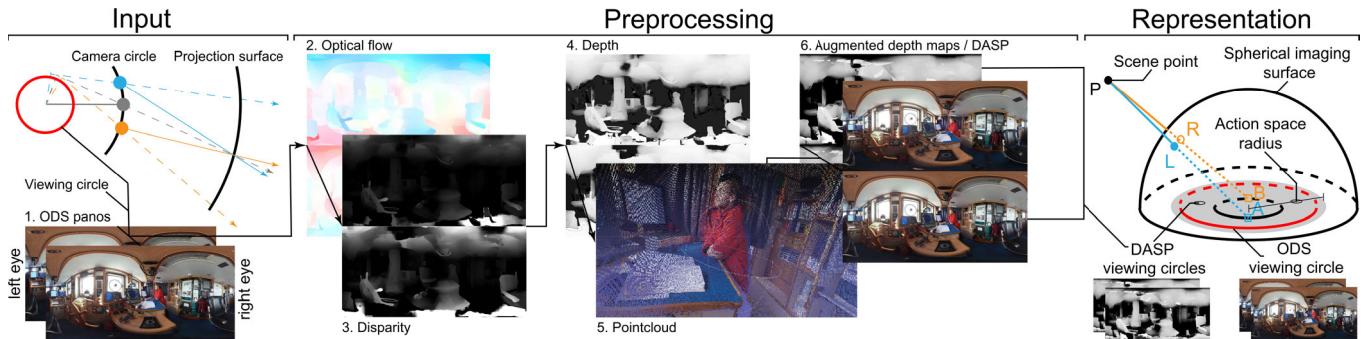


Figure 1: We create a 6-DoF VR experience from a single omnidirectional stereo (ODS) pair of a scene. Our approach takes an ODS panorama as input (1), along with the radius of the viewing circle. We determine disparities between the left and right eye views using optical flow (2). The disparities (3) are used to obtain depth per pixel (4) creating a pointcloud (5) used to generate a DASP.

## ABSTRACT

We present an end-to-end pipeline that enables head-motion parallax for omnidirectional stereo (ODS) panoramas. Based on an ODS panorama containing a left and right eye view, our method estimates dense horizontal disparity fields between the stereo image pair. From this, we calculate a depth augmented stereo panorama (DASP) by explicitly reconstructing the scene geometry from the viewing circle corresponding to the ODS representation. The generated DASP representation supports motion parallax within the ODS viewing circle. Our approach operates directly on existing ODS panoramas. The experiments indicate the robustness and versatility of our approach on multiple real-world ODS panoramas.

## 1 INTRODUCTION

The recent technological evolution of head-mounted displays allows for experiencing new types of digital media at previously unprecedented levels of immersion [3]. While rendering synthetic geometry only requires few adjustments, providing real-world content in virtual reality (VR) remains a challenging research problem.

Omnidirectional stereo (ODS) [2, 4] is the de-facto industry standard for real-world VR experiences. The sense of depth from the left and right eye images allows for a higher level of immersion, but the lack of motion parallax causes head motion to feel unnatural.

Recently, depth augmented stereo panoramas (DASP) [6] were introduced to enable head motion parallax in spherical panoramas. Utilising additional depth information, novel views can be synthesised with head-motion parallax.

We take a hybrid approach by first estimating implicit geometry using optical flow, and then utilising the encoded 3D pointcloud of the DASP approach [6] for novel-view synthesis. Our approach takes as input an ODS image and estimates a pointcloud by using disparities between both views to create a depth augmented stereo

panorama. This enables 6-DoF head motion during rendering without storing any explicit geometry besides one augmented depth map per eye.

## 2 METHOD

The input to our approach is an ODS panorama pair to be transformed into a DASP, along with the interpupillary distance (IPD) which determines the viewing circle. As a first step, we estimate dense correspondences within the ODS panorama. According to the IPD and correspondences, we calculate the depth to lift every pixel up and form a sparse 3D pointcloud. The last step is backprojecting the reconstructed pointcloud into the DASP representation, forming our augmented depth maps. Along with the ODS pair, this opens up the possibility of novel-view synthesis and therefore head-motion parallax at runtime.

We use  $\theta \in [-\pi, \pi]$  as the azimuth angle and  $\varphi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$  as the elevation angle. We assume that  $\theta = -\pi$  corresponds to the left edge and  $\varphi = \frac{\pi}{2}$  to the top edge of the ODS panoramas.

**Disparity estimation** Corresponding scene points in ODS footage project to the same horizontal scanline and all disparities share the same direction. Due to the stitching process for ODS synthesis, these are not always satisfied though. Hence, we widen the search space using a state-of-the-art optical flow approach [1] instead of using restrictive disparity estimators working on 1D lines.

To account for the actual ODS representation we filter the resulting optical flow  $F = \{F_x, F_y\}$  and keep only the horizontal component as disparity. This results in two dense disparity maps, one for the left and one for the right view. Hereby, the disparity  $d(\mathbf{x})$  of a pixel  $\mathbf{x}$  corresponds to the angular disparity  $\Delta\theta$ .

**Depth estimation** Given the estimated disparity maps, we compute depth from the right and left viewing circles,  $D_R$  and  $D_L$ .

$$D = \frac{r}{\sin\left(\frac{\Delta\theta}{2}\right)} \quad \text{and thus} \quad D_{R/L} = \sqrt{D^2 - r^2}. \quad (1)$$

Here,  $r$  is the radius of the viewing circle (half the IPD) and  $\Delta\theta$  the angular disparity. We incorporate the corrected elevation angle  $\varphi_{R/L}$ :

$$\hat{D}_{R/L} = \frac{D_{R/L}}{\cos(\varphi_{R/L})}. \quad (2)$$

<sup>\*</sup>Both authors contributed equally

<sup>†</sup>e-mail: T.B.Bertel@bath.ac.uk

<sup>‡</sup>e-mail: muehlhausen@cg.cs.tu-bs.de

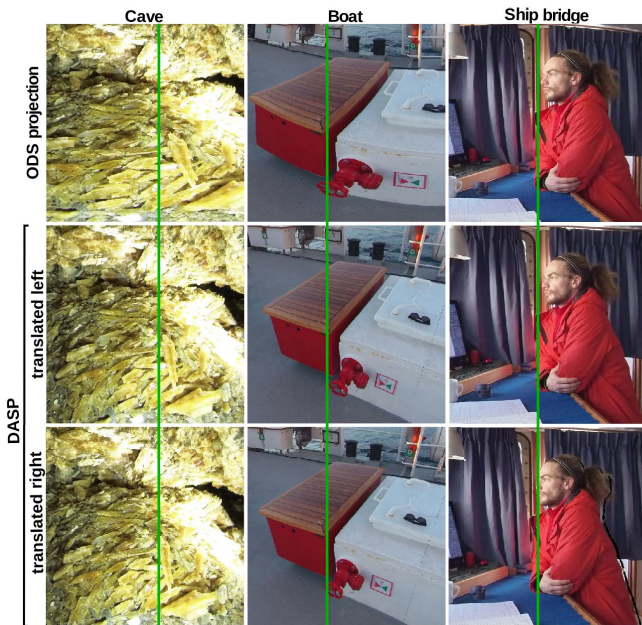


Figure 2: Comparison of ODS and DASP scene representations and results for a range of real-world scenes. The ODS projection only allows for head rotation without motion parallax. The DASP representation combines omnidirectional panoramas of the ODS projection with the depth information of the pointcloud, allowing for head motion parallax within a small motion-range.

As the perceivable disparity gets close to zero for large depth values, we clamp our values  $\hat{D}_{R/L}$  to a predefined maximum of 50 m for an IPD of 6.4 cm. At this point, we have augmented our input panorama pair in the ODS representation with a depth map for each view, hence resulting in a DASP representation.

**Depth augmented stereo panoramas (DASP)** DASPs [6] compactly represent a 3D scene centered around an ODS viewing circle. The quality of the representation strongly depends on the available pointcloud and thus on the estimated disparities. We choose a small action space (1 cm), in which we achieve high-quality motion parallax within the viewing circle. Exceeding the viewing circle imposes view extrapolation which is difficult to solve in general and even harder if geometry for new unseen scene volumes is unknown. We use Thatte et al.’s strategy for rendering the DASP [6]. This results in more immersive VR experiences based on ODS by additionally providing head-motion parallax.

### 3 RESULTS AND DISCUSSION

We create DASPs for ODS images captured and stitched using the Insta360 Pro camera rig and show qualitative results for a set of ODS datasets. All images are rendered as a monocular view using the DASP representation and its associated view synthesis procedure. For easier analysis of occurring rendering holes, we did not perform hypothesis merging [6]. We define the predefined motion range as the radius of the viewing circle plus 1 cm (compare to the gray disk in Fig. 1, Representation) and set the imaging sphere radius to 1.9 m in all our experiments. In our prototype, our pipeline creates a DASP from a high-resolution ODS panorama pair (7680×3840 pixels) and pre-computed optical flow end-to-end in about ten minutes using a desktop PC with an Intel i7 quad core CPU and 32 GB RAM.

Since we estimate the pointcloud, and thus depth, only by using the ODS panoramas, our reconstructed DASPs do not support much head motion range because of the needed view extrapolation. Nevertheless, by applying our pipeline to a standard ODS, we can obtain a DASP which produces appealing head-motion parallax within the

available viewing circle. We show results for head translations within the viewing circle in Fig. 2. Firstly, by performing image-based rendering with pointclouds, the DASP representation produces results with correct perspective. Secondly, we translate through the viewing disk from left to right (compare ODS projection and left and right views of the DASP in Fig. 2). Observe the scene content moving away and towards the green vertical line respectively. The supported head-motion is limited due to the small motion-range chosen to avoid excessive view extrapolation.

Our pipeline to create 6-DoF VR experiences from consumer ODS is mainly limited in two aspects: (1) the quality of the estimated disparities, and (2) the inherent quality of the resulting DASP and its own limitations. The pointcloud reconstructed from the ODS viewing circle is sufficiently dense to produce high-quality results, but it is not complete. The attempt to look into occluded scene volumes, which are not geometrically represented within the DASP representation, will thus fail.

### 4 CONCLUSION

ODS panoramas are constrained by a single viewing circle allowing only for head rotations around a fixed point while only providing correct binocular disparity when looking along the equator. The DASP representation allows for full-head motion parallax in 6-DoF.

We have shown a way of transforming real-world VR content in form of ODS into a DASP and shown results from several real-world datasets. The quality of our reconstructed DASP is sufficient to provide viewpoints with full 6-DoF head-motion parallax within a small motion range. The results indicate that our method can enrich existing ODS footage in order to create and provide more exciting real-world VR applications.

**Future work** We would like to evaluate our pipeline and the produced DASP representation in more detail. The quality of the DASP must be assessed in an interactive VR application by conducting a user study, and parameters, such as the predefined motion range and the imaging sphere radius, need to be examined. Furthermore, it would be interesting to apply our pipeline to ODS data created from different sources. We finally would like to investigate ways of empowering the DASP representation by using a layered approach close in spirit to the idea of layered depth images [5] and look into options to enhance the quality of the reconstructed pointcloud, i.e. by 3D inpainting techniques.

### ACKNOWLEDGMENTS

This work was funded by the German Science Foundation (DFG MA2555/15-1 “Immersive Digital Reality” and DFG INST 188/409-1 FUGG “ICG Dome”), EU H2020 MSCA FIRE (665992), EPSRC-CDE (EP/L016540/1), RCUK grant CAMERA (EP/M023281/1), a UKRI Innovation Fellowship (EP/S001050/1), a Rabin Ezra Scholarship, and an NVIDIA Corporation GPU Grant.

### REFERENCES

- [1] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *CVPR*, 2017. doi: 10.1109/CVPR.2017.179
- [2] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *TPAMI*, 14(2):257–262, February 1992. doi: 10.1109/34.121792
- [3] G. A. Koulouris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt. Near-eye display and tracking technologies for virtual and augmented reality. *Comput. Graph. Forum*, 38(2):493–519, May 2019. doi: 10.1111/cgf.13654
- [4] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic stereo imaging. *TPAMI*, 23(3):279–290, 2001. doi: 10.1109/34.910880
- [5] J. Shade, S. Gortler, L. wei He, and R. Szeliski. Layered depth images. In *SIGGRAPH*, pp. 231–242, July 1998. doi: 10.1145/280814.280882
- [6] J. Thatte, J.-B. Boin, H. Lakshman, G. Wetzstein, and B. Girod. Depth augmented stereo panorama for cinematic virtual reality with focus cues. In *ICIP*, pp. 1569–1573, 2016. doi: 10.1109/ICIP.2016.7532622