Omnidirectional Stereo

Christian Richardt
University of Bath

Synonyms

Omnistereo, ODS

Related Concepts

Calibration of a Non-single Viewpoint System
Image Stitching
Omnidirectional Camera
Video Mosaicing

Definition

Omnidirectional stereo (ODS) is a type of multi-perspective projection that captures horizontal parallax tangential to a viewing circle. This data allows the creation of stereo panoramas that provide plausible stereo views in all viewing directions on the equatorial plane.

Background

The term 'omnidirectional stereo' was first coined by Ishiguro et al. [1] in 1990, who used it in the context of autonomous mapping and exploration of an unknown environment. Their approach places a video camera on a rotating arm that is driven by a stepper motor (see Figure 1). They then take vertical slit images from the sensor image for creating the left/right stereo panoramic views. As their primary goal is to map an environment, they use ODS images for depth estimation of scene points rather than display. They also present a binocular stereo method for depth estimation from two ODS images with a measure for direction-dependent uncertainty.

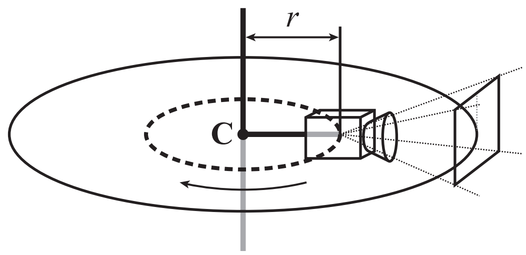

Figure 1: Omnidirectional stereo images can be captured with a single camera mounted on an arm of length r that rotates about a point $\mathbf{C}$. Figure adapted from Ishiguro et al. [1].

Before the term 'omnidirectional stereo' became established with its current meaning, it was used more broadly for any stereo imaging system, which captures stereo panoramas with disparity, regardless of direction. Gluckman et al. [2], for example, created a real-time system consisting of two catadioptric cameras that are vertically displaced. These cameras capture omnidirectional views of an environment that therefore primarily differ by vertical disparity. While this works just as well for depth estimation, these vertical stereo panoramas are unsuitable for being viewed by humans. This is because our eyes are horizontally displaced and the human visual system thus expect horizontal disparity, not vertical disparity. In the following, the focus therefore lies on stereo panoramas with horizontal disparity.

Since the early days of ODS for mapping and robotics, the main application has shifted towards display on monitors, projection screens and head-mounted virtual reality displays: Peleg et al. [3] popularized 'omnistereo' panoramas with automatic disparity control, Richardt et al. [4] proposed improvements for creating high-quality, high-resolution stereo panoramas, and Anderson et al. [5] and Schroers et al. [6] developed state-of-the-art systems for capture and display of real-world virtual reality video. The benefit of ODS video is that time-varying stereo panoramas can easily be packed into a traditional video, which can be processed, stored and transmitted on existing video streaming platforms just like any other video. This has established omnidirectional stereo as a widely supported format by video streaming pipelines and virtual reality displays.

Theory

Omnidirectional stereo is fundamentally a multi-perspective projection that is created from rays that are tangential to a *viewing circle* [3]. This projection can, for example, be captured by a slit camera that moves along a circular path and looks in the tangential direction. The slit images captured at different positions on the viewing circle are mosaicked into a panorama, producing the ODS projection. The two possible tangential directions give rise to the left and right panoramas, as illustrated in Figure 2. The diameter of the viewing circle is usually chosen to be the average human interpupillary distance of 65 mm; however, other sizes are possible.

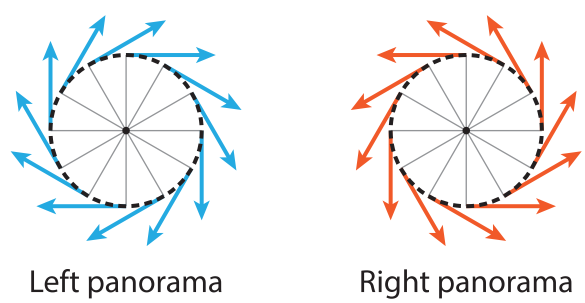

Figure 2: Omnidirectional stereo projection creates two panoramas, for the left and right views, using rays tangential to the viewing circle.

As we shall see, most ODS systems do not capture the ODS projection directly. The only implementation of a pair of rotating slit cameras is by Konrad et al. [7]. The key benefit of this approach is that the system directly captures the ODS projection without requiring computationally expensive processing. In addition, this system natively supports scenes containing nearby objects, occlusions, thin and repetitive textures, transparent and translucent surfaces, and specular and refractive objects, which remain challenging for other capture approaches. However, the prototype camera was limited to only 5 frames per second due to the fast rotation of the camera assembly that is required for operation.

Ishiguro et al. [1] first described and Peleg et al. [3] later popularized a single-camera approach that rotates a standard perspective camera on a larger *capture circle* that is concentric to the viewing circle. The camera needs to rotate a full 360 degrees to capture a complete ODS panorama, which assumes a static scene during the capture time. As illustrated in Figure 3, the camera can simultaneously capture rays for both the left and the right stereo panorama. While this works perfectly in the 2D equatorial plane, out-of-plane rays introduce vertical distortion [5], and perspective cameras introduce perspective distortion. Richardt et al. [4] proposed techniques for correcting perspective distortion and for removing stitching artifacts using flow-based blending. Stabilization steps were also introduced that reduce drift from image alignment and enable hand-held capture, where the camera's pose often deviates from the ideal position and orientation along the capture circle.

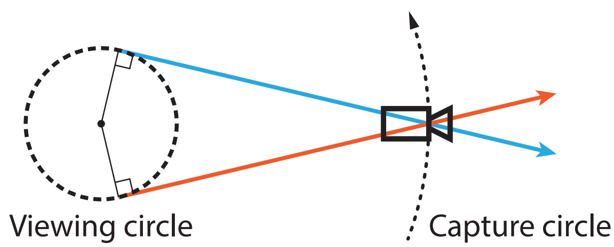

Figure 3: A single rotating camera can capture all rays tangential to the viewing circle, by rotating it on a larger circle. Note that this only works perfectly within the plane containing the viewing and capture circles; other rays are captured with vertical distortion. Figure adapted from Anderson et al. [5].

However, the rotating-camera approach is not practical for capturing ODS *videos*, as the camera needs to rotate a full 360 degrees for each video frame. For example, this would require very fast 30 revolutions per second for capturing ODS video with 30 frames per second. Peleg et al. [3] explored theoretical optical designs involving spiral mirrors and spiral lenses that would enable single-shot ODS capture and thus single-camera ODS video. However, these designs were never realized. The TORNADO system by Tanaka and Tachi [8] was the first practical implementation based on these designs. Elongated prism sheets deflect the light rays entering the cylindrical camera system and emulate 32 slit cameras that rotate over time. Inside the cylinder, a hyperboloidal mirror reflects light rays into a stationary video camera. The projections of the slit cameras into the video camera are separated in two different ways: (1) based on their spatial location in images, or (2) using linear/circular polarizers. Aggarwal et al. [9] proposed a different approach that uses a carefully designed mirror with a 'coffee filter' shape and reflects multiple slit camera images into the video camera without any additional optical elements. High-resolution simulations validate this concept, but manufacturing inaccuracies severely reduce the visual quality of real-world captures. Low visual quality and image resolutions are problems shared by all current catadioptric single-camera approaches [8, 9].

Approaches with multiple omnidirectional cameras improve some of the quality problems caused by complex optical elements. Chapdelaine-Couture and Roy [10] presented an approach for three or more fisheye cameras that are arranged in a regular polygon, with parallel viewing directions towards the sky. They observed that the planes spanned by the optical axes of adjacent pairs of cameras can be used for cutting their omnidirectional images and stitching them together with minimal disparity mismatch. However, this approach can only capture the upper hemisphere of any scene. Matzen et al. [11] proposed a system that uses two consumer spherical cameras, which have two 180-degree fisheye lenses each. The cameras are mounted side-by-side with a distance of 64 mm and the captured imagery is sliced by the plane going through the cameras (orthogonal to the optical axes) and recombined to create the necessary left/right panoramas for ODS imagery. Their approach cancels vertical disparity at the seams using warping in image space, and uses this disparity to correct horizontal disparity. As horizontal disparity is maximal in front of the cameras but zero along the camera baseline, this correction depends on the azimuth angle. In both cases, the resolution of the resulting ODS video was not sufficient for high-quality virtual reality experiences due to limited sensor resolutions [10, 11].

The highest visual quality and image resolution of ODS video has been demonstrated by approaches that use multiple video cameras spaced evenly on a circle, like a discretized version of the continuously rotating camera of earlier approaches [1, 3, 4]. Anderson et al. [5] pack 16 GoPro cameras with fisheye optics into a tighly packed ring of 28 cm diameter, demonstrating that this is a sweet spot between more cameras and a smaller ring diameter that reduces vertical distortion. Their work comprises a detailed analysis of the sources of vertical parallax and distortion, and presents a flow-based stitching pipeline that produces temporally coherent ODS video, as available on YouTube, for example. Schroers et al. [6] also use 16 video cameras, but formulate their video pipeline based on continuous light field reconstruction from sparse samples. This naturally leads to a panoramic image formation model that allows computer-generated imagery to be composited with captured real-world content. They also provide an analysis of the minimum visible depth that can be stitched or reconstructed, based on the number of cameras in the rig and their field of view.

Application

The initial application for omnidirectional stereo was for mapping and exploring environments using robots [1]. In this scenario, a single rotating camera provides a low-cost and effective solution for reconstructing a given static scene. Ishiguro et al. [1] demonstrate how to estimate the depth of points from a single ODS image but also from a binocular pair of ODS images. They further show that the accuracy of depth estimation depends on the direction relative to the camera baseline, with parallel directions having the highest uncertainty. This uncertainty can be decreased by integrating multiple observations in the form of local maps into a single global map.

Since then, omnidirectional stereo has become the de-facto standard projection for stereo panoramas that are used for virtual reality photography and video. Google's Cardboard Camera app, for example, is based on the principles described by Richardt et al. [4] for creating high-quality, high-resolution stereo panoramas. Multiple independently developed systems for VR videos—including Google Jump [5], Schroers et al. [6] and Facebook Surround360 [12]—have arrived at remarkably similar hardware: multi-view camera rigs with 14–16 cameras that are packed into a tight ring to capture omnidirectional stereo video. As of 2019, this defined the state of the art in widely available virtual reality imagery and video.

References

[1] Hiroshi Ishiguro, Masashi Yamamoto, and Saburo Tsuji. Omnidirectional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):257–262, February 1992. ISSN 0162-8828. doi: 10.1109/34.121792.

[2] Joshua Gluckman, Shree K. Nayar, and Keith J. Thoresz. Real-time

omnidirectional and panoramic stereo. In *Proceedings of the Image Understanding Workshop*, 1998.

[3] Shmuel Peleg, Moshe Ben-Ezra, and Yael Pritch. Omnistereo: Panoramic stereo imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):279–290, 2001. ISSN 0162-8828. doi: 10.1109/34.910880.

[4] Christian Richardt, Yael Pritch, Henning Zimmer, and Alexander Sorkine-Hornung. Megastereo: Constructing high-resolution stereo panoramas. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1256–1263, June 2013. doi: 10.1109/CVPR.2013.166.

[5] Robert Anderson, David Gallup, Jonathan T. Barron, Janne Kontkanen, Noah Snavely, Carlos Hernandez, Sameer Agarwal, and Steven M. Seitz. Jump: Virtual reality video. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 35(6):198:1–13, November 2016. ISSN 0730-0301. doi: 10.1145/2980179.2980257.

[6] Christopher Schroers, Jean-Charles Bazin, and Alexander Sorkine-Hornung. An omnistereoscopic video pipeline for capture and display of real-world VR. *ACM Transactions on Graphics*, 37(3):37:1–13, August 2018. ISSN 0730-0301. doi: 10.1145/3225150.

[7] Robert Konrad, Donald G. Dansereau, Aniq Masood, and Gordon Wetzstein. SpinVR: Towards live-streaming 3D virtual reality video. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 36(6): 209:1–12, November 2017. ISSN 0730-0301. doi: 10.1145/3130800.3130836.

[8] Kenjie Tanaka and Susumu Tachi. TORNADO: omnistereo video imaging with rotating optics. *IEEE Transactions on Visualization and Computer Graphics*, 11(6):614–625, November 2005. ISSN 1077-2626. doi: 10.1109/TVCG.2005.107.

[9] Rajat Aggarwal, Amrisha Vohra, and Anoop M. Namboodiri. Panoramic stereo videos with a single camera. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3755–3763, June 2016. doi: 10.1109/CVPR.2016.408.

[10] Vincent Chapdelaine-Couture and Sébastien Roy. The omnipolar camera: A new approach to stereo immersive capture. In *Proceedings of the International Conference on Computational Photography (ICCP)*, April 2013. doi: 10.1109/ICCPhot.2013.6528311.

[11] Kevin Matzen, Michael F. Cohen, Bryce Evans, Johannes Kopf, and Richard Szeliski. Low-cost 360 stereo photography and video capture. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, 36(4): 148:1–12, July 2017. ISSN 0730-0301. doi: 10.1145/3072959.3073645.

[12] Facebook. Surround 360 instruction manual. GitHub, July 2016. URL https://github.com/facebook/Surround360/blob/master/surround360_design/assembly_guide/Surround360_Manual.pdf.