

Spherical Light Fields

Bernd Krolla¹

<http://av.dfki.de/~krolla>

Maximilian Diebold²

<http://hci.iwr.uni-heidelberg.de/Staff/mdiebold/>

Bastian Goldlücke³

<http://www.informatik.uni-konstanz.de/cvia/>

Didier Stricker¹

<http://av.dfki.de/stricker>

¹ German Research Center

for Artificial Intelligence,
Trippstadter Straße 122,
Kaiserslautern, Germany

² Heidelberg Collaboratory

for Image Processing,
Speyerer Straße 6,
Heidelberg, Germany

³ Department for Computer and

Information Science,
University of Konstanz,
Konstanz, Germany

Abstract

A full view spherical camera exploits its extended field of view to map the complete environment onto a 2D image plane. Thus, with a single shot, it delivers a lot more information about the surroundings than one can gather with a normal perspective or plenoptic camera, which are commonly used in light field imaging. However, in contrast to a light field camera, a spherical camera does not capture directional information about the incident light, and thus a single shot from a spherical camera is not sufficient to reconstruct 3D scene geometry.

In this paper, we introduce a method combining spherical imaging with the light field approach. To obtain 3D information with a spherical camera, we capture several independent spherical images by applying a constant vertical offset between the camera positions and combine the images in a *Spherical Light Field* (SLF). We can then compute disparity maps by structure tensor orientation analysis on epipolar plane images, which in this context are 2D cuts through the spherical light field with constant azimuth angle. This method competes with the acquisition range of laser scanners and allows for a fast and extensive recording of a given scene. We benchmark our approach by comparing disparity maps of ray-traced scenes against its ground truth. Further we provide disparity maps of real world datasets.

1 Introduction

Since projects such as Microsoft *Street Side* [18] or Google *Street View* [2] provide numerous spherical images to online users, spherical imaging experienced increasing attention in the recent past. To acquire such spherical images, a wide variety of hardware devices is available, delivering results of varying quality and accuracy. The devices divide up into professional solutions [24, 20, 22, 29] and consumer oriented camera devices such as [2, 24]. Torii *et al.* provide a fundamental and elegant definition of spherical cameras in [25], subsuming

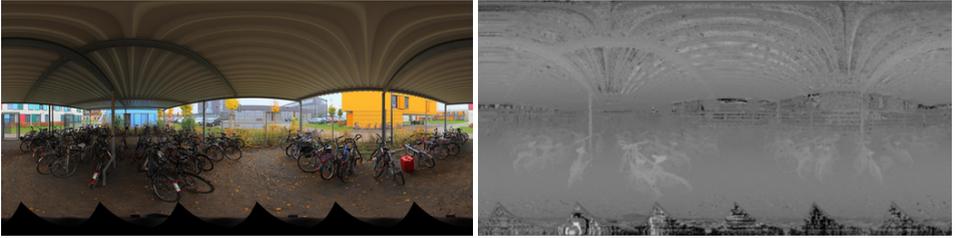


Figure 1: Left: Central image of the captured Spherical Light Field. Right: The corresponding disparity map.

central dioptric and catadioptric cameras under the assumption of known camera parameters into this camera model

Spherical cameras are able to handle interesting application scenarios, which are not realizable with standard perspective cameras. Pagani *et al.* [14] researched *Structure from Motion* approaches using full spherical cameras, whereas the work of Aly and Bouquet [1] is more focused on the calibration of unordered sets of spherical images. Gutierrez *et al.* [10] showed that visual SLAM can be performed without loss of image features caused by camera rotation when using spherical instead of perspective cameras. Furthermore, as application oriented approaches, 3D reconstruction using multi spherical stereo has been employed to reconstruct the 3D environment of an static scene [6, 11].

The combination of omnidirectional images with *High Dynamic Range* (HDR) imaging as introduced in [14, 21] expands image processing possibilities such as noise reduction, shadow handling or avoidance of under- and over-exposed image regions.

Aside from spherical image acquisition, light field imaging is another area of research which has gained more and more attention over the last few years, in particular with the advent of consumer oriented light field cameras [8, 9, 12, 27]. Since the 4D light field implicitly encodes 3D information, it is easily possible to robustly compute disparity maps from the captured data. Recently, orientation analysis on epipolar plane images using the structure tensor has been identified as an efficient way to compute disparity maps over the complete light field [28].

Related work. The interface between light field imaging and omnidirectional camera systems has been addressed by recent research, whereas to the best of our knowledge full spherical images have not been considered. Birklbauer and Bimber created panorama light fields by stitching multiple perspective light fields taken by a rotated light field camera [4]. Due to the used devices [9, 12] remains the vertical field of view (FOV) limited.

Taguchi *et al.* [23] used an array of spherical mirrors to model catadioptric cameras for wide angle light field rendering. While providing dense depth estimation and refocusing capabilities for the captured scenes, the setup entails decreasing tangential resolution close to the mirror borders, limiting the FOV to $150^\circ \times 150^\circ$.

Unger *et al.* [26] employed a similar capturing configuration than in [23] as well as a fisheye-camera translated on a plane to capture hemispherical HDR images of a scene. Aiming at the rendering of artificial objects in the captured environment, the total acquisition time takes up to 12 hours for a single scene. This restricts the application scenario to indoor environments, since constant illumination conditions during the capturing are crucial for the subsequent light-field processing.

An alternative approach to obtain full spherical depth and disparity maps for a surround-

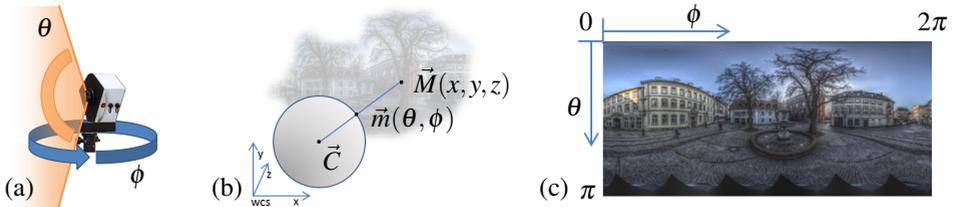


Figure 2: (a) Spherical image acquisition using a rotating tripod mounted camera equipped with a fish eye lens. (b) The image results from the back projection of 3D points $M(x, y, z)$ to their corresponding image points $m(\theta, \phi)$ assuming C to be the cameras center of projection. (c) In the current work, the resulting image is a *High Dynamic Range* (HDR) image with a resolution of 14000×7000 pixel and is parametrized using spherical coordinates $\phi[0, 2\pi]$ and $\theta[0, \pi]$.

ing scene is to use laser scanners, which measure the depth in an bounded range around the device. Even though these achieve highly accurate reconstructions, they are in general very costly devices and the resulting scans are commonly provided without texture information of the scene.

Contributions. In this work, we combine spherical imaging with light field analysis and introduce the concept of Spherical Light Field recording. To capture a *Spherical Light Field* (SLF), we obtain several spherical images from different altitudes, which results in a 3D data structure parametrized by the two angular directions and the height of the capturing device. We show that by computing 2D cuts through this structure with fixed azimuth angle, we obtain the analogon of an epipolar plane image, where we can efficiently perform depth reconstruction via orientation analysis. Furthermore, this makes it possible to directly adapt light field analysis techniques which rely on epipolar plane image analysis to the scenario of omnidirectional scene acquisition. Compared to conventional light field cameras, we acquire significantly more information about the surrounding scene. In particular, we also capture the scene in high dynamic range. In the context of this work, we can thus benefit especially from improved texture representation as well as improved illumination estimation to increase the performance of subsequent analysis of the SLF. We demonstrate that SLF offers the possibility of very short acquisition time using small sets of 9-13 high-resolution spherical images for disparity estimation.

2 Spherical Light Field disparity estimation

2.1 Spherical image acquisition

Our proposed approach for SLF acquisition relies on the utilization of spherical cameras as shown in Figure 2(a). A convenient description of this camera type is provided by Torii *et al.* [23], who consider a spherical camera to consist of a camera center C with a surrounding unit sphere acting as projection surface. This definition implies that no intrinsic parameters such as focal length or distortion values known from perspective imaging need to be considered.

According to the collinearity constraint, any 3D point M of the camera’s environment is

mapped through the camera center C to its corresponding image point m , see Figure 2(b). Any position within the resulting spherical image is uniquely defined by the image coordinates $\phi \in [0, 2\pi)$ and $\theta \in [0, \pi)$. By applying the Mercator projection [9], the spherical image is conformally mapped to an image on a cylinder surface Π , see Figure 2(c).

Note that this kind of data representation implies a significant distortion of image content close to the image poles ($\theta \rightarrow 0$ and $\theta \rightarrow \pi$). However, it assures that any content of the scene is shifted along the latitude-axis of the image, with respect to vertical displacement of the camera position. Therefore, this representation is suitable to allow epipolar plane image (EPI) reconstruction, as outlined in the following section.

2.2 Spherical Light Fields

To describe a SLF, we define a new parametrization for the camera domain and the surrounding spherical 2D mapped image, see Figure 3(a). We take the cylinder surface Π and

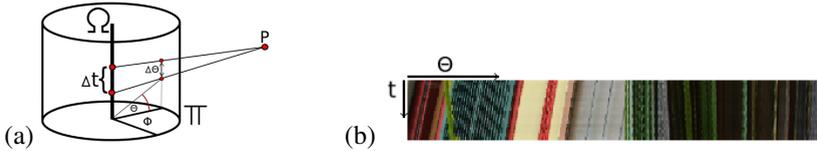


Figure 3: (a) Parametrisation of the Spherical Light Field. (b) Example of an *Epipolar Plane Image (EPI)* assembled from 15 images.

denote the center line with Ω . The cylinder surface Π is parametrized by the image coordinates $(\phi, \theta) \in \Pi$. The line Ω contains the focal points $t \in \Omega$ of all possible camera positions in vertical direction.

A Spherical Light Field can now be described by a function

$$L : \Omega \times \Pi \rightarrow \mathbb{R} \quad (t, \phi, \theta) \mapsto L(t, \phi, \theta), \quad (1)$$

where $L(t, \phi, \theta)$ defines the intensity of the incident light ray on the image plane (ϕ, θ) passing through the focal point t .

To estimate the disparity, we address a 2D slice Σ_{ϕ^*} of the SLF by setting ϕ to a fixed value ϕ^* . The restriction of the light field to such a slice is called an epipolar plane image (EPI), and is formally defined as

$$S_{\phi^*} : \Sigma_{\phi^*} \rightarrow \mathbb{R} \quad (2)$$

$$(\theta, t) \mapsto S_{\phi^*}(\theta, t) := L(t, \phi^*, \theta). \quad (3)$$

An example is shown in Figure 3(b).

Assuming a Lambertian scene, the EPI yields information about the disparity of a scene point in the form of orientated lines. Each line corresponds to the projection of a scene point, and its slope is directly related to the parallax, so is in a one-to-one correspondence to the distance of this point from the camera center.

To compute the disparity on the EPI, we can thus perform an orientation analysis on the given EPI S_{ϕ^*} , using the structure tensor

$$J = \tau * \begin{pmatrix} (S_{\theta})^2 & S_{\theta} S_t \\ S_t S_{\theta} & (S_t)^2 \end{pmatrix} =: \begin{pmatrix} J_{\theta\theta} & J_{\theta t} \\ J_{\theta t} & J_{tt} \end{pmatrix} \quad (4)$$

with the abbreviations

$$S_t := \sigma * \frac{\partial S}{\partial t}, \quad S_{\theta} := \sigma * \frac{\partial S}{\partial \theta}. \quad (5)$$

The orientation angle and thus the disparity map d for the EPI S_{ϕ^*} can be computed directly from the components of the structure tensor via

$$d = \tan \left(\frac{1}{2} \arctan \left(\frac{J_{tt} - J_{\theta\theta}}{J_{\theta t}} \right) \right). \quad (6)$$

As a reliability measure of the estimated disparity, one can employ the coherence κ defined by

$$\kappa = \sqrt{\frac{(J_{tt} - J_{\theta\theta})^2 + 4J_{\theta t}^2}{(J_{tt} + J_{\theta\theta})^2}}. \quad (7)$$

The full set of disparity and coherence maps, exemplary shown in Figure 6, is computed by iterating over all EPIs from the SLF and storing the computed disparity and coherence values at the corresponding azimuthal slice.

3 Spherical Light Field acquisition

3.1 Synthetic data

Using the Blender [B] software, we designed a synthetic scene, which was used to validate the disparity estimation results of the proposed SLF approach. The required spherical images are assembled column-wise. Each column is extracted from the rendered image of a perspective camera, where the camera image plane loops around a common center of projection located on Ω . Here, we use 360 perspective camera images with a resolution of $20px \times 3000px$ to generate a spherical image with a resulting size of $7200px \times 3000px$. Two of the different scenes we use are shown in Figure 6(a).

Using these datasets, we will compare the estimated depth maps to ground truth depth maps, and validate the reliability measure κ by comparing to the actual estimation error.

3.2 Real data

To acquire spherical images in real environments, we use the omnidirectional dioptric *Civetta* camera manufactured by Weiss AG [W], see Figure 2(a). This camera is equipped with a fish eye lens and provides omnidirectional $360^\circ \times 180^\circ$ HDR images by stitching multiple perspective images together.

Since the camera software handles distortion and overlaps of the input images, the resulting spherical HDR images comply with the spherical camera model introduced previously.

By applying the Mercator projection, they are mapped to a plane and stored as EXR-files [13] with a resolution of 14000×7000 pixels, see Figure 2(c). The file size of up to 320MB results from a combination of high resolution and a 24bit HDR color representation.

For the capturing process, we need to consider that the camera requires a static scene to provide optimal results. Since the HDR characteristic of the images is obtained by capturing multiple images with varying exposure time from the same position, moving objects cause artifacts in the resulting image.

To acquire the actual SLF, camera positions of increasing height were engaged by varying the tripod’s height by a fixed amount on the order of several millimeters. Besides the desired pixel offset along the latitude coordinate θ , also minor offsets along the longitude coordinate ϕ occurred due to the manual adjustment of the tripod height. Thus, to assure optimal data quality for a reliable EPI generation, a realignment of the images was performed as a first post processing step after the image capturing.

To perform the image realignment along the ϕ coordinate, standard computer vision approaches were applied by extracting and matching SIFT-features [14] from the different images. To improve the robustness of the realignment, feature extraction was limited to a strip along the image equator ($\pm 60^\circ$ latitude) by masking distorted image regions close to the image poles. After rejecting outlying matches, the averaged offset between the images could be retrieved up to subpixel-precision and was used to align the captured set of spherical images against each other.

4 Results

To assess the reliability of geometry reconstruction with the proposed SLF approach, we evaluate the disparity estimation on synthetic and real world data. We generate two test sequences and store ground truth depth for each captured image in the spherical light field using Blender. For the captured real world data, we generate disparity maps.

4.1 Synthetic data evaluation

For the reliability evaluation, we compute the pixel mean absolute error of the computed disparity map versus the ground truth data, and compute a difference map by subtracting the ground truth data from the estimated disparities, as seen in Figure 6(d). The center view of the ray-traced scenes can be observed in Figure 6(a). The setup has a spherical camera to camera baseline of $5mm$ in vertical direction. The resulting disparity map of the scene matches to the ground truth data quite well, see Figure 6(b) and Table 1. In the synthetic scene, some computational artifacts appear at the bottom of the disparity map. These artifacts only appear in synthetic data, and are caused by disparity values which are outside the range where the structure tensor orientation analysis can yield reliable results (± 2 pixels).

Inside the measurable range, we are able to compute the disparity of a scene with very high accuracy. Furthermore, the coherence map is a very good indicator of which pixels are valid. To see this, we binarize the coherence map, and mark pixels which have a coherence of less than 0.03 as “unreliable”. Second, we also threshold the error map and mark all pixels with an error of more than 0.2 pixels as “wrongly estimated”. The comparison of both binary maps shows that over 99% of pixels are wrongly estimated if and only if the estimate is unreliable.

Finally, as numeric quality estimate, we compute both the mean absolute error for the entire image (MAE_E) and the mean absolute error for the reliable pixels (MAE_R). Results can be seen in Table 1.

Dataset (synthetic)	Disparity range [px]	MAE_E [px]	MAE_R [px]
1	3.000 – 9.000	1.869	0.074
2	5.000 – 11.000	3.875	1.581

Table 1: Disparity range, mean absolute error of the entire image (MAE_E) and reliably estimated pixels (MAE_R) for the two ray-traced scenes.

4.2 Real data evaluation

To further evaluate the proposed SLF approach, we record multiple datasets of real scenes. Key parameters of the different scenes, which include indoor and outdoor scenarios, are provided in Table 2. The acquisition time scales linearly with the number of captured images, while the capturing of a single spherical image takes roughly one minute with the camera used in this work.

In all scenarios, the HDR information we recorded was exploited to provide input image sets with characteristics optimal for the disparity estimation software. In particular, the two tone-mapping algorithms *pfstmo_reinhard02* and *pfstmo_mantiuk06* [14] were applied to maximize image contrast for the processing step. Figure 4 shows examples for tone-mapped image closeups as well as the corresponding disparity estimates.

The disparity maps are estimated on a workstation equipped with an Intel® Core™i7 CPU at 2.80GHz and 16GB memory. Processing times listed in Table 2 are obtained by running the estimation software on a single core. Due to hardware memory limitations, the captured spherical images had to be rescaled to a resolution of 2357×1179 pixels. Resulting disparity maps with the corresponding center and view coherence maps are shown in Figure 5 and provided in more detail as supplementary material.

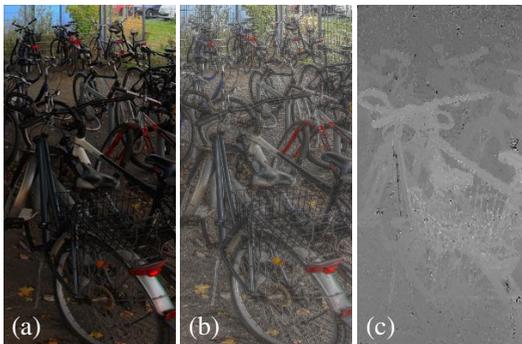


Figure 4: Closeups of the recorded bike scene. (a) Tonemapped input image. (b) Input image with optimized contrast for SLF processing. (c) The resulting disparity map.

Dataset (real)	Images	Baseline	Kernel size [px]	Processing time [s]
<i>Bikes</i>	11	5mm	11×7	41
<i>Basement</i>	13	5mm	13×13	54
<i>Courtyard</i>	19	5mm	15×7	75

Table 2: Three datasets of captured real-world scenes with varying number of spherical images, but constant camera to camera baseline of 5mm. Due to memory limitations, all datasets were scaled to a resolution of 2357×1179 for disparity estimation.

5 Conclusion

In this work, we capture spherical light fields in synthetic and real-world environments using full spherical cameras. The mapping of the resulting spherical images to a conformal representation on a 2D plane allows to easily construct epipolar plane images, on which it is possible to apply orientation analysis for fast and accurate disparity estimation. The resulting full view spherical disparity maps can then be employed for a 3D scene reconstruction of the camera’s surroundings. Benchmarks on synthetic datasets demonstrate good accordance with the ground truth data. Furthermore, combining spherical and HDR imaging approaches for the capturing of real scenes can greatly simplify the task of disparity estimation due to e.g. improved contrast.

In conclusion, the concept of spherical light fields presents a promising avenue to expand the applications for light field processing, in particular towards those which require a detailed and complete map of the surroundings.

For further evaluations besides the datasets contained in this document, the reference may be made to the supplementary material submitted to this conference.

Acknowledgements

This work was funded by Sony Deutschland, Stuttgart Technology Center, EuTEC and is a result of the research cooperation between STC, EuTEC, the Heidelberg Collaboratory for Image Processing (HCI) and the German Research Center for Artificial Intelligence (DFKI).

We would like to thank Thimo Emmerich, Yalcin Incesu and Oliver Erdler from STC, EuTEC for their feedback to this work and all the fruitful discussions.

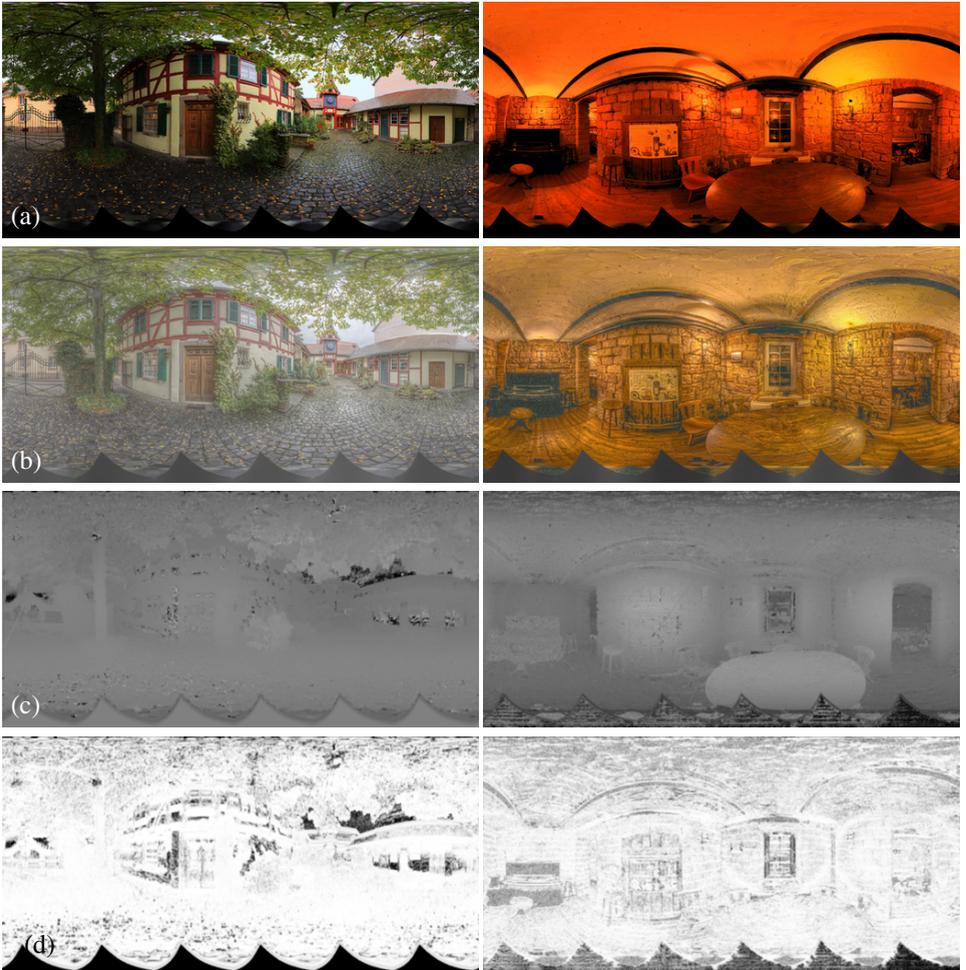


Figure 5: Results of the *courtyard* and *basement* dataset. The center view images of the datasets are shown with realistic (a) and contrast-optimized tonemapping (b). The disparity maps, computed with the structure tensor approach (c), and the coherence maps, indicating the reliability of the disparity estimation (d), are shown below.

References

- [1] Mohamed Aly and J-Y Bouguet. Street view goes indoors: Automatic pose estimation from uncalibrated unordered spherical panoramas. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 1–8, 2012.
- [2] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver. Google Street View: Capturing the world at street level. *IEEE Computer*, 43(6):32–38, 2010.
- [3] Mats Bentsen, Geir Evensen, Helge Drange, and Alistair Jenkins. Coordinate transformation on a sphere using conformal mapping. *Monthly Weather Review*, (127): 2733–2740, 1999.
- [4] Clemens Birklbauer and Oliver Bimber. Panorama light-field imaging. In *SIGGRAPH Posters*, page 61, 2012.
- [5] Blender Foundation. Blender, 2014. URL <http://www.blender.org/>.
- [6] Alan Brunton, Jochen Lang, and Eric Dubois. Efficient multi-scale stereo of high-resolution planar and spherical images. In *2nd International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 120–127, 2012.
- [7] Ricoh Company. Ricoh theta, 2013. URL <http://theta360.com/en/>.
- [8] Todor Georgiev, Zhan Yu, Andrew Lumsdaine, and Sergio Goma. Lytro camera technology: theory, algorithms, performance analysis. In *IS&T/SPIE Electronic Imaging*, volume 8667, pages 1J1–1J10. International Society for Optics and Photonics, 2013.
- [9] Raytrix GmbH. Raytrix, 2014. URL <http://www.raytrix.de/>.
- [10] Daniel Gutierrez, Alejandro Rituerto, JMM Montiel, and José Jesús Guerrero. Adapting a real-time monocular visual SLAM from conventional to omnidirectional cameras. In *IEEE International Conference on Computer Vision Workshops*, pages 343–350, 2011.
- [11] Kim Hansung and Adrian Hilton. 3D Modelling of Static Environments Using Multiple Spherical Stereo. In *Proc. European Conference on Computer Vision*, page 2, 2010.
- [12] Lytro Inc. Lytro, 2014. URL <https://store.lytro.com/>.
- [13] Industrial Light and Magic. OpenEXR fileformat, 2013. URL <http://www.openexr.com/>.
- [14] LizardQ GbR. LizardQ, 2014. URL <http://www.lizardq.com/>.
- [15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [16] Steve Mann. Compositing multiple pictures of the same scene. In *Proceedings of the 46th Annual IS&T Conference*, volume 2, 1993.

- [17] Rafał Mantiuk, Grzegorz Krawczyk, Radosław Mantiuk, and Hans-Peter Seidel. High dynamic range imaging pipeline: Perception-motivated representation of visual content. In *Human Vision and Electronic Imaging XII*, volume 6492 of *Proceedings of SPIE*, San Jose, USA, 2007. SPIE. URL <http://pfstools.sourceforge.net/>.
- [18] Microsoft. Microsoft Streetside. URL <http://www.microsoft.com/maps/streetside.aspx>.
- [19] Alain Pagani and Didier Stricker. Structure from motion using full spherical panoramic cameras. In *Proceedings of the Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (OMNIVIS)*, volume 11, 2011.
- [20] Pointgrey. Ladybug 5, 2014. URL <http://ww2.ptgrey.com/Products/Ladybug5/>.
- [21] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: Acquisition, display, and image-based lighting: 2nd ed.* Elsevier, 2010.
- [22] SpheronVR AG. SceneCam[®] HDR, 2014. URL <https://www.spheron.com/products/visual-asset-management.html>.
- [23] Yuichi Taguchi, Amit Agrawal, Ashok Veeraraghavan, Srikumar Ramalingam, and Ramesh Raskar. Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *ACM Transactions on Graphics-TOG*, 29(6):172, 2010.
- [24] Tamaggo Inc. Ibi 360, 2014. URL <http://ibi360.tamaggo.com/>.
- [25] Akihiko Torii, Atsushi Imiya, and Naoya Ohnishi. Two- and three-view geometry for spherical cameras. In *Proceedings of the sixth workshop on omnidirectional vision, camera networks and non-classical cameras*. Citeseer, 2005.
- [26] Jonas Unger, Andreas Wenger, Tim Hawkins, Andrew Gardner, and Paul Debevec. Capturing and rendering with incident light fields. In *Proceedings of the 14th Eurographics workshop on Rendering*, pages 141–149. Eurographics Association, 2003.
- [27] Kartik Venkataraman, Dan Lelescu, Jacques Duparré, Andrew McMahon, Gabriel Molina, Priyam Chatterjee, Robert Mullis, and Shree Nayar. Picam: an ultra-thin high performance monolithic camera array. *ACM Transactions on Graphics (TOG)*, 32(6):166, 2013. URL <http://www.pelicanimaging.com/technology/>.
- [28] S. Wanner and B. Goldluecke. Variational Light Field Analysis for Disparity Estimation and Super-Resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.
- [29] Weiss, AG. Civetta 360° Camera. URL <http://www.weiss-ag.org/solutions/civetta/>.