2016 16th International Conference on Control, Automation and Systems (ICCAS 2016)
Oct. 16–19, 2016 in HICO, Gyeongju, Korea

3D Reconstruction of Structures using Spherical Cameras with Small Motion

Sarthak Pathak, Alessandro Moro, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama

Department of Precision Engineering, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Email: {author-last-name}@robot.t.u-tokyo.ac.jp

Abstract: In this research, a method for dense 3D reconstruction of structures from small motion of a spherical camera is proposed. Spherical cameras can capture information from all directions enabling measurement of the entire surrounding structure at once. The proposed technique uses two spherical images clicked at slightly displaced positions near the structure, followed by a combination of feature-point matching and dense optical flow. Feature-point matching between two images alone is usually not accurate to give a dense point cloud because of outliers. Moreover, calculation of the epipolar direction with feature point matching is susceptible to noise with small displacements. However, spherical cameras have unique parallax properties allowing use of dense, global information. Taking advantage of this, the global, dense optical flow field is used. The epipolar geometry is densely optimized based on the optical flow field for an accurate 3D reconstruction. A possible use of this research could be to measure large infrastructures (bridges, tunnels, etc.) with minimal robot motion.

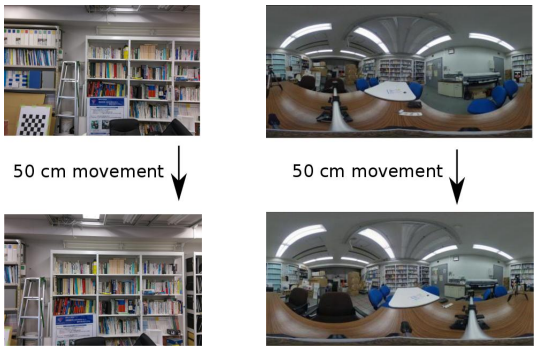
Keywords: 3D Reconstruction, Spherical Images, Optical Flow

1. INTRODUCTION

Spherical cameras (such as the Ricoh Theta S) that can capture real-time information from all directions are quite useful in robotics. One basic task that is often performed by robots is dense 3D reconstruction. Especially, 3D reconstructions of large infrastructures such as bridges is important for their maintenance and inspection [3]. Image based 3D reconstruction relies on camera motion and the direction and length of motion are critical for its accuracy. In many cases, it can be difficult or costly to move the robot in a desired manner or desired length to capture a large number of images. Hence, this research proposes a method that makes it possible to reconstruct a structure from only two spherical images clicked near it with a small displacement between them.

It is well-known that a spherical field of view has less rotation and translation ambiguities as compared to perspective images [12], [2]. However, for small displacements, sparse point correspondences alone can lead to a noisy epipolar direction that cannot be used for a dense reconstruction. This is because of two reasons - the presence of outliers, and poor conditioning of the epipolar constraint over small displacements. Even methods like RANSAC [4] may miss a few points because of the weak point-to-line epipolar constraint. These issues can be hazardous for attempting dense reconstruction using two images alone.

Fortunately, spherical cameras have helpful parallax properties. If a perspective camera moves, there can be a significant change in the image, especially if it is close to a structure. The same movement for spherical cameras will create a much lesser change due to the large field of view (Figure 1). This can allow use of dense, global information. It has been shown how dense optical flow can help in accurate epipolar estimation in perspective cameras [16]. In this paper, the same concept is extended to take advantage of spherical image geometry. The optical flow field is optimized to a purely translational state based on a non-linear minimization, converting the two images into a rectified stereo pair for dense 3D reconstruction.



(a)Perspective Camera Movement of 50 cm

(b)Spherical Camera Movement of 50 cm

Fig. 1 Spherical images[2] undergo much lesser change than perspective cameras for the same movement, allowing for dense, global algorithms

1.1 Overview

The proposed approach uses two spherical images with a small displacement and arbitrary rotation between them. 8-point RANSAC-based sparse correspondences are used for an initial estimate and to bring the images to the same orientation, enabling computation of dense optical flow. The 5 degree of freedom epipolar geometry parameters are then optimized over the dense optical flow to align all flow vectors along their epipolar lines. Fi-

[2]Note: In this paper, all spherical images are displayed in the equirectangular projection unless otherwise mentioned.

nally, the length of the aligned optical flow vectors along epipolar lines is directly converted to a dense reconstruction. The originality of this approach lies in combining the advantages of spherical image parallax properties with dense optical flow methods to accurately estimate the epipolar geometry and consequently a dense reconstruction of the surrounding structure.

1.2 Related Research

Previous research in spherical image-based 3D reconstruction largely involves spherical stereo vision [8], [7]. [8] uses calibrated camera positions as input and [7] works under mechanical vertical alignment of the two cameras. There are also pure sparse correspondence-based methods like [9] and [14] which (as mentioned earlier) can lead to ambiguous epipolar geometry estimates with small displacements. To solve these ambiguities, some approaches use multiple images followed by global bundle adjustment [13]. However, it is not suitable for use with a small number of (or only two) images.

Instead of sparse point-based methods, methods like [11] which optimize a dense global error function can be much more accurate and applicable to spherical images. On similar lines, pixel-to-pixel displacements can be expressed with dense optical flow instead. Considering the applicability of dense methods on spherical images, an optical flow based pipeline for epipolar estimation and dense reconstruction is proposed. In another work as well, estimation of spherical camera rotation based on dense optical flow based minimization [15] was proposed by us. In this work, it is extended to obtain the full epipolar geometry estimate and a 3D reconstruction.

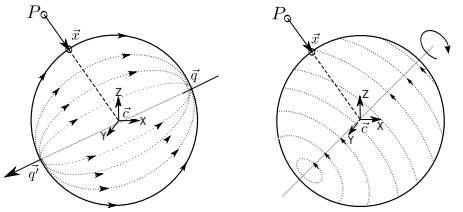
2. SPHERICAL IMAGE EPIPOLAR GEOMETRY

Spherical image epipolar geometry, as the name suggests, is based on the surface of a unit sphere. Hence, all the pixels are represented as unit radius vectors $\vec{x} = \begin{bmatrix} x & y & z \end{bmatrix}^t$. From one view to another, corresponding points follow the well-known epipolar geometry constraint:

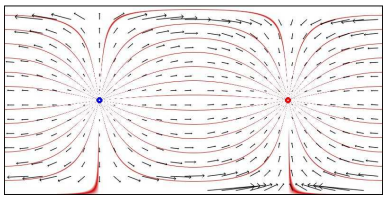
$$\mathbf{x'^t E x} = \mathbf{0} \tag{1}$$

where $\mathbf{E}$ is the essential matrix that contains the rotation and translation information, and $\mathbf{x}$ and $\mathbf{x}'$ are the corresponding matched points in the two images.

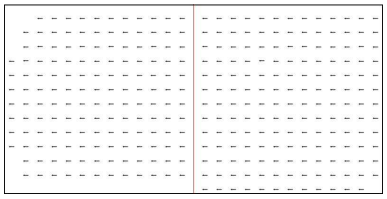
As explained earlier, in case of a small movement, dense optical flow can be calculated on the spherical image. Optical flow exhibits certain interesting patterns on the surface of a sphere that makes it possible to uniquely distinguish a translational and rotational flow. [12] first discussed these patterns and suggested a theoretical searching based approach to estimate the rotation followed by another for translation. Essentially, a purely translational flow involves the vectors 'emerging' from a point and 'converging' at another point following great circles on the surface of the sphere. These points



(a)Translation (spherical) (b)Rotation (spherical)



(c)Translation (Equirectangular Projection). The blue and red points indicate the epipoles while the red lines are epipolar lines. The motion field is aligned along them.



(d)Rotation (Equirectangular Projection) The axis of rotation is in the vertical direction, while the arrows indicates direction.

Fig. 2 Motion Fields on the Unit Sphere for the camera undergoing (a) Pure Translation (b) Pure Rotation in spherical and equirectangular projections.

are nothing but the epipoles and the great circles are the epipolar lines. Meanwhile, a purely rotational flow involves optical flow vectors forming loops around the axis of rotation. Since any camera motion consists of translation and rotation, the optical flow pattern is always a superimposition of the translational and rotational patterns. Both are shown in Figures 2.

Since spherical images contain information from all directions, an important property emerges. After an initial computation of optical flow between two images, they can be rotated to any desired orientation and the optical flow state at that orientation can be reprojected. Further, if an image is rotated in a way to cancel out its rotation with respect to another, the resultant optical flow should be purely translational in nature. Thus, the rotation as well as translation direction can be estimated. The proposed technique for epipolar geometry retrieval is dependent on this concept of 'derotation'.

3. EPIPOLAR GEOMETRY ESTIMATION BY DEROTATION

The proposed method involves derotating one of the spherical images to search a point at which the optical

flow becomes purely translational. For this, a suitable error measure to indicate deviation from a translational state needs to be defined. The property of a translational state of motion, as seen from Figure 2 is that all the optical flow vectors are aligned along the epipolar lines, pointing towards the epipole. Thus, deviations of the optical flow vectors from their epipolar lines are chosen to form the error function. This is similar to the non-linear refinement steps of typical structure from motion algorithm [5]. The difference being that it is not based on sparse point matches that can be affected by outlier noise. The regularization property of dense optical flow ensures that there are no local outliers that can drastically affect the result, as thoroughly explained in [16].

To represent rotation, the three euler angles α, β, and γ in the $x-y-z$ notation are used. As for translation, its direction is represented as one of the epipoles in spherical coordinates as θ and ϕ. Thus, $G(\alpha,\beta,\gamma,\theta,\phi)$ is defined as the parameter vector that uniquely defines the epipolar geometry of the two images. Note that all parameters are taken in radians. A derotation-based minimization is formulated to optimize this parameter vector in order to bring the two images to a purely translational state. However, the search space can be quite large as each parameter can take any angular value. Thus, an appropriate initialization technique is necessary, as described in the next subsection.

3.1 Initialization

As mentioned earlier, the aim is to build a 3D model from small, arbitrary movements. Thus, the displacement of the image due to translation alone is not expected to be large. (Moreover, spherical images do not change much with translation.) Thus, optical flow can be estimated in case of pure translation. However, if the rotation is too large, the images will be quite displaced and optical flow cannot be calculated. In order to solve this problem and to provide an initial value for the minimization-based epipolar estimation, the process can be initialized using the sparse feature correspondences.

Many sparse correspondences in both images are computed using A-KAZE features [1], which work well under distortions. They are filtered in an 8-point RANSAC approach [6][4] to estimate an inlier set and the epipolar matrix $\mathbf{E}$. Using singular value decomposition, $\mathbf{E}$ is decomposed into the rotation matrix and the translation vector. The translation vector forms the initial value for the translation part of the parameter vector. The images are derotated using the estimated rotation matrix. Now, the orientations are approximately the same, displaced by the inaccuracy of the point matching based approach. Hence, the dense optical flow-based refinement is done using an initial value of zero rotation.

3.2 Minimization

Now, the refinement of epipolar geometry with a least-squares minimization will be explained. The initial state is that of zero rotation, as calculated from the noisy point-correspondence estimate. Now, it is optimized to ensure perfect zero rotation and an optimal translation direction using the following error function. In each iteration, the spherical image (denoted by S) is derotated with the estimates of α, β, and γ and the dense optical flow field between the two images is calculated. At every pixel $\vec{x}$, the optical flow vector is denoted by $\vec{f}$, relative to $\vec{x}$. The error function is defined based on how close this field is to a pure translational state. Thus, point $\vec{x}$ moves to point $\vec{x}+\vec{f}$. $\vec{f}$ consists of the rotational flow component $\vec{f_{rot}}$ and the translational flow component $\vec{f_{trans}}$. In a pure translational state, there should be no rotation and $\vec{f}$ should be only translational. Mathematically,

$$\begin{aligned} \vec{f_{trans}} &= \vec{f} \\ \vec{f_{rot}} &= 0 \end{aligned} \tag{2}$$

Eq.(2) describes the condition for a purely translational flow field. If the parameter vector $G(\alpha,\beta,\gamma,\theta,\phi)$ is optimized to make sure Eq.(2) is satisfied as much as possible at all points in S, the images can be brought to a translational state. Thus, the difference between $\vec{f}$ and $\vec{f_{trans}}$ at every pixel should be minimized on the spherical image. Since the magnitude of $\vec{f_{trans}}$ depends on the yet unknown depth, the angular difference between $\vec{f}$ and $\vec{f_{trans}}$ can be taken as the error at point $\vec{x}$, denoted by Ω. Ω can be calculated by finding the deviation of $\vec{f}$ from the epipolar great circle C_q defined by the epipole $\vec{q}$ and point $\vec{x}$ (as shown in Figure 3):

$$\Omega = \left(\arccos\left(\frac{(\vec{q}\times\vec{x})\cdot\left((\vec{x}+\vec{f})\times\vec{x}\right)}{|(\vec{q}\times\vec{x})|\times|\left((\vec{x}+\vec{f})\times\vec{x}\right)|} \right) \right) \tag{3}$$

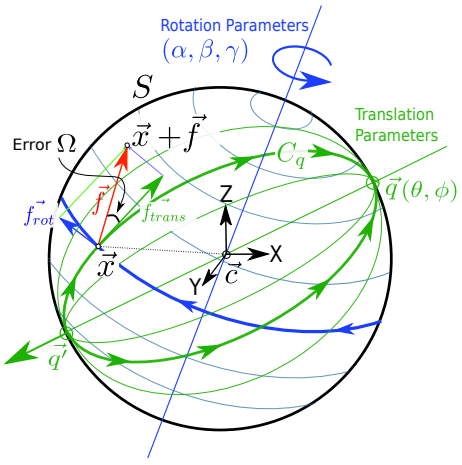


Fig. 3 Error at $\vec{x}$. ($\vec{f}$ - red, $\vec{f_{trans}}$ - green). To find angle Ω, the cross products $(\vec{x}+\vec{f})\times\vec{x}$ and $\vec{q}\times\vec{x}$ are taken in order to define the great circles along $\vec{f_{trans}}$ and $\vec{f}$ and find the angle between them.

A dense non-linear least squares error for each $\vec{x}\in S$ is defined and the optimal $G(\alpha,\beta,\gamma,\theta,\phi)$ that aligns all

flow vectors along epipolar lines is found:

$$\underset{G(\alpha,\beta,\gamma,\theta,\phi)}{\text{minimize}} \sum_{\forall \vec{x} \in S} \Omega^2 \tag{4}$$

These types of non-linear least squares problems are easily solved using the popular Levenberg-Marquardt approach [10]. Once the optimal parameter vector is determined, the resultant arrangement is two densely rectified stereo images with only translation between them which is suitable for 3D reconstruction, as described in the next section.

4. 3D RECONSTRUCTION

At the end of the optimization, a purely translational state of optical flow is obtained. In other words, the images have been stereo-rectified. The magnitude of each optical flow vector $\vec{f}$ at $\vec{x}$ is now directly related to its inverse depth, depending on its angular distance from the epipole $\vec{q}$. Thus, its 3D position can now be triangulated in the following manner.

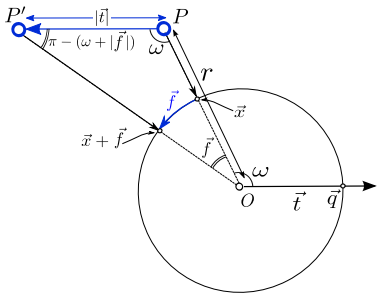


Fig. 4 Triangulation

As seen in Figure 4, ω is the angular distance from epipolar point $\vec{q}$ and $|\vec{f}|$ is the angular disparity (magnitude of $\vec{f}$ divided by the radius $= 1$). The law of sines in triangle POP' gives the radius r of point P:

$$r = |\vec{t}| \times \frac{\sin(\omega + |\vec{f}|)}{\sin(|\vec{f}|)} \tag{5}$$

The scale of the translation vector $\vec{t}$ is arbitrarily taken as 1 without loss of generality to give the final structure.

5. EXPERIMENTS

The algorithm was implemented in C++ and the recently developed DeepFlow [17] algorithm was used to estimate the optical flow. An experimental setup was created in order to check the quality of the 3D reconstruction. Two textured planar boards were placed perpendicular to each other (Figure 5). The popular Ricoh Theta S spherical camera was used. A spherical image was captured at an original position. Then, the camera was displaced upwards by 5 cm and rotated arbitrarily to capture another. The image pair is shown in Figure 6. An initial

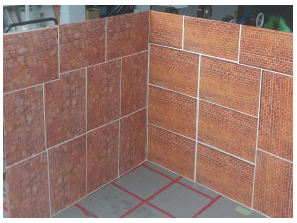


Fig. 5 Experimental Setup for 3D Reconstruction

(a)Image clicked at original position

(b)After 5 cm translation and arbitrary rotation

Fig. 6 Spherical Stereo Image Pair of the setup

estimate from A-KAZE feature points [1] was performed, as shown in Figure 7.

The proposed dense refinement was applied to this pair of images. Figure 8 shows the results. Figure 8 (a) shows the original optical flow field obtained with the initial estimate of epipolar positions (red and blue are $\vec{q}$ and $\vec{q'}$) from the point matching. The optical flow vectors are misaligned from the epipolar lines drawn in red. Figure 8 (b) shows the final translational optical flow, with aligned optical flow vectors and final epipoles. Stereo disparity along the epipolar lines was estimated from the final optical flow state (Figure 8 (d)) and the 3D structure was reconstructed as mentioned in the previous section. For comparison, reconstruction using the epipolar geometry estimated only by the sparse points was also performed. Both reconstructions are shown in Figure 9. One perspective view and the top view are shown for comparison. The reconstruction based only on point matching is quite distorted, while the one after the proposed refinement resembles the actual structure.

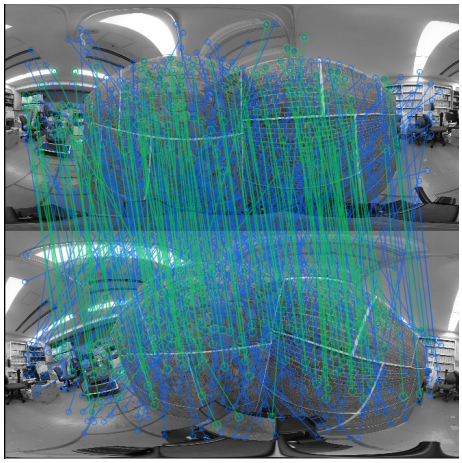
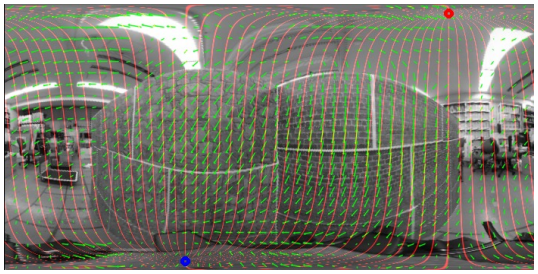
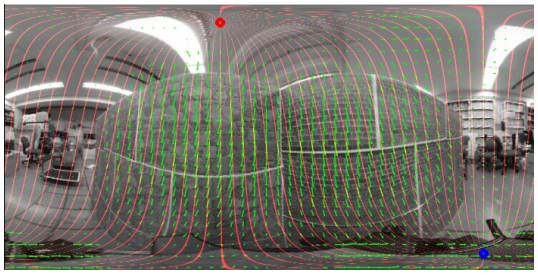


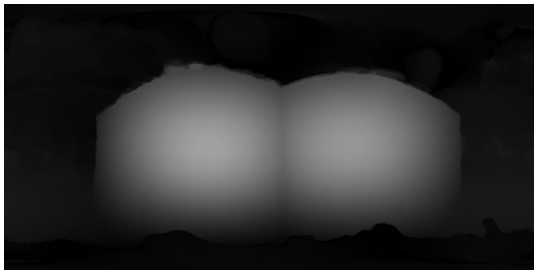
Fig. 7 RANSAC with the 8-point algorithm for initial estimation. All point matches are in blue, while the inliers are in green.



(a)Before proposed refinement, using only sparse point matching: Initial optical flow and epipolar state with unaligned optical flow vectors.

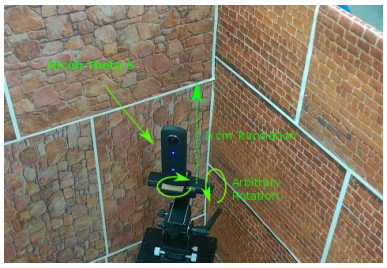


(b)After proposed refinement: Final, refined optical flow and epipolar state with optical flow vectors aligned along epipolar lines (red): Translational Flow

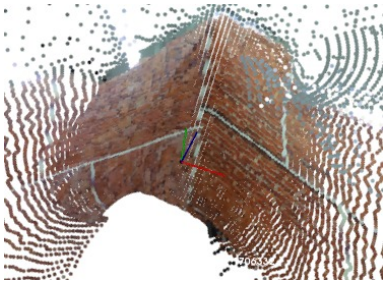


(c)Disparity magnitude along epipolar curves

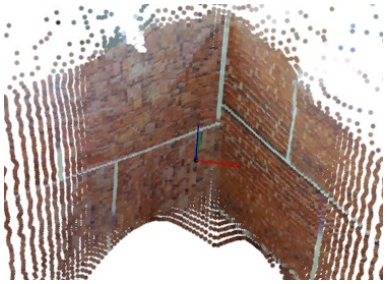
Fig. 8 Output of the algorithm applied to the image pair



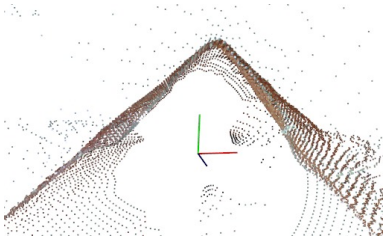
(a)Experiment Setup



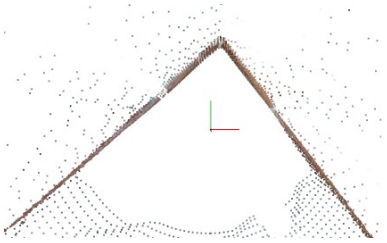
(b)Distorted reconstructed view from sparse point matching only



(c)Accurate reconstructed view after proposed optical flow-based refinement



(d)Distorted reconstructed top view from sparse point matching only



(e)Accurate reconstructed top view after proposed optical flow-based refinement

Fig. 9 3D Reconstructed Views and camera positions before and after the proposed refinement. A cleaner, more accurate structure can be noticed after the refinement.

6. CONCLUSION

In this research, a method for dense reconstruction method based on capturing spherical images and using an optical flow-based refinement was proposed. This enables 3D reconstruction of structures over a wide angle with a small displacement. Considering the advantages of dense optical flow-based epipolar estimation suggested by [16], it was extended it to spherical images. Two unique spherical image properties were taken advantage of: parallax properties allowing for dense global algorithms, and the large field of view, allowing for measurement of a large area.

This research is intended as a precursor to a video-based structure from motion approach involving large surrounding structures with a spherical camera placed on a mobile robot. The main purpose was to solve the problem of robot movement in such scenarios. In future work, the stereo disparity estimation algorithm will be improved and this refinement will be used to formulate a multi-view 3D reconstruction approach, similar to [13].

Acknowledgement

This work was in part supported by the 'Cross-ministerial Strategic Innovation Promotion Program (SIP), Infrastructure Maintenance, Renovation, and Management', Council for Science Technology and Innovation (funding agency: NEDO).

REFERENCES

[1] P. F. Alcantarilla, J. Nuevo, and A. Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *Proceedings of the British Machine Vision Conference*, September 2013.

[2] T. Brodsky, C. Fermüller, and Y. Aloimonos. Directions of motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 26(1):5–24, January 1998.

[3] H. Fathi, F. Dai, and M. Lourakis. Automated as-built 3d reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. *Advanced Engineering Informatics*, 29(2):149 – 161, April 2015.

[4] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[5] J. Fujiki, A. Torii, and S. Akaho. Epipolar geometry via rectification of spherical images. In *Proceedings of the Third International Conference on Computer Vision/Computer Graphics Collaboration Techniques*, pages 461–471, March 2007.

[6] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997.

[7] H. Kim and A. Hilton. 3d scene reconstruction from multiple spherical stereo pairs. *International Journal of Computer Vision*, 104(1):94–116, August 2013.

[8] N. Kita. Dense 3d measurement of the near surroundings by fisheye stereo. In *Proceedings of the IAPR Conference on Machine Vision Applications, 2011*, pages 148–151, June 2011.

[9] Z. Kukelova, J. Heller, M. Bujnak, A. Fitzgibbon, and T. Pajdla. Efficient solution to the epipolar geometry for radially distorted cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2309–2317, December 2015.

[10] M.I.A. Lourakis. levmar: Levenberg-marquardt nonlinear least squares algorithms in c/c++. `http://www.ics.forth.gr/~lourakis/levmar/`, July 2004.

[11] A. Makadia and K. Daniilidis. Rotation recovery from spherical images without correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1170–1175, August 2006.

[12] R. C. Nelson and J. Aloimonos. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58(4):261–273, March 1988.

[13] A. Pagani and D. Stricker. Structure from motion using full spherical panoramic cameras. In *Proceedings of the IEEE International Conference on Computer Vision (Workshops)*, pages 375–382, November 2011.

[14] D. V. Papadimitriou and T. J. Dennis. Epipolar line estimation and rectification for stereo image pairs. *IEEE Transactions on Image Processing*, 5(4):672–676, April 1996.

[15] S. Pathak, A. Moro, A. Yamashita, and H. Asama. A decoupled virtual camera using spherical optical flow. In *Proceedings of the IEEE International Conference on Image Processing (accepted)*, September 2016.

[16] L. Valgaerts, A. Bruhn, M. Mainberger, and J. Weickert. Dense versus sparse approaches for estimating the fundamental matrix. *International Journal of Computer Vision*, 96:212–234, January 2012.

[17] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. DeepFlow: Large displacement optical flow with deep matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1385 – 1392, December 2013.