Exploring the Limits

Applying State-of-the-Art Stereo Matching Algorithms to Rectified Ultra-Wide Stereo

Slezák, Filip; Laursen, Morten S.; Moeslund, Thomas B.

Published in:
Proceedings - 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2024

DOI (link to publication from Publisher):
10.1109/CVPRW63382.2024.00141

Publication date:
2024

Document Version
Accepted author manuscript, peer reviewed version

Exploring the Limits: Applying State-of-the-Art Stereo Matching Algorithms to Rectified Ultra-Wide Stereo

Filip Slezák
AGCO A/S, Aalborg University, DK
filip.slezak@agcocorp.com

Morten S. Laursen
AGCO A/S, DK
mortenstigaard.laursen@agcocorp.com

Thomas B. Moeslund
Aalborg University, DK
tbm@create.aau.dk

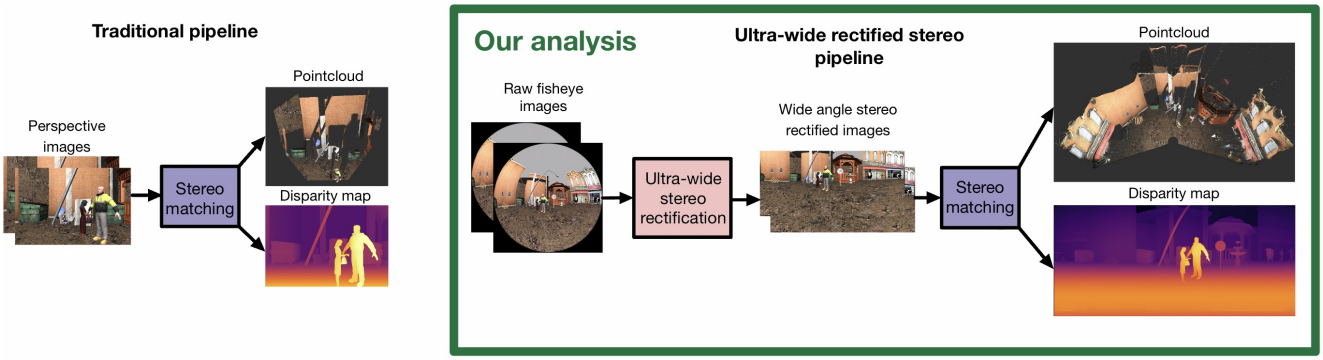

Figure 1. Our work analyzes the potential of reusing state-of-the-art stereo-matching algorithms to achieve ultra-wide depth perception.

Abstract

Stereo cameras leveraging two-view geometry have predominantly focused on narrow field-of-view rectified stereo using the pinhole camera model. This research trend overlooks the complexities and potential of wide-angle stereo systems, which necessitate the use of wide-angle fisheye optics that can not be well approximated by pinhole camera model. Consequently, a lack of standardized form leads researchers to explore various strategies. Currently, a dichotomy exists between utilizing raw images directly or rectifying them. Wide-angle stereo rectification opens the potential to reuse the latest state-of-the-art (SOTA) algorithms designed for pinhole rectified stereo as a black box. However, rectification comes at the cost of severe distortions throughout the image and non-linear triangulation of 3D structure. The literature currently lacks a thorough examination of the implications of these distortions and the impact of applying the latest SOTA algorithms to stereo-rectified wide-angle images. Our work addresses this gap by conducting an exhaustive analysis of the wide-angle rectified stereo framework, delivering concrete recommendations for developing accurate wide-angle stereo systems.

1. Introduction

Stereo cameras are a popular research topic for estimating the structure of the environment. The traditional setup involves two images captured at slightly different locations, inducing visual parallax, where the same point in 3D is projected to a slightly different position in the images. The task of estimating the structure of the environment is analogous to estimating a pixel-wise displacement, also referred to as a disparity map.

Earlier works such as SGM [11] and ELLAS [7] have relied on traditional methods and achieved remarkable success but struggle with occlusions and texture-less regions. Recently, deep learning-based methods have become dominant in the field, predominantly utilizing the RAFT architecture [30]. These approaches have produced algorithms for estimating disparity that are not only highly accurate but also demonstrate robustness when dealing with thin structures, occlusions and texture-less surfaces.

Despite the maturity of the field and strong momentum on popular benchmarks [28][21], there is a fundamental limitation. The primary assumption is the form of rectified stereo using perspective images, which are fundamentally limited by the field-of-view (FOV) they can effectively represent. At wide angles approaching 180°, the repre-

sentation becomes inefficient causing, compression in the middle and stretching around the sides.
One solution is to utilize optics better suited for efficiently representing a wide FOV, such as fish-eye, which can capture even up to 220° FOV. However, the non-linear projection model causes the epipolar lines to no longer project into straight lines, greatly complicating the design of correspondence search algorithms and prohibiting the direct re-use of advanced disparity estimation algorithms designed for rectified stereo as a black box.

Non-linearity introduces distortions generally considered problematic, motivating various methods to combat their effects. These include searching in raw images directly along the curved epipolar lines [23][25], plane/sphere sweeping [9][22] or designing custom representations which minimize local distortions [12].
The additional complexity introduced by more complicated problem formulation has compromised the quality of the corresponding finding robustness, falling behind methods from the rectified stereo state-of-the-art. Even the most recent methods still use block matching to compute correspondences, resulting in considerable speckling [12].
An alternative method involves adjusting wide-angle images to align the epipolar lines with the image rows, allowing for re-using advanced rectified stereo-matching algorithms. However, the wide-angle stereo rectification comes at the cost of non-homogenous resampling of the source image, resulting in varying image quality and localized distortions.

Reusing state-of-the-art stereo-matching algorithms for wide-angle stereoscopic distance estimation is a greatly under-studied problem. Schneider [29] has analyzed the effects of triangulation on local distance estimation accuracy but only used a simplified one-dimensional model relative to the incident angle. Beekmans [2] has assessed the accuracy of estimating distances in relation to a flat surface; however, this approach does not represent an omnidirectional perception framework.
However, the error propagation from disparity to distance estimates is only one aspect to consider. No work currently analyzes the interaction between wide-angle rectified images and state-of-the-art algorithms designed for perspective images.
We introduce a detailed analysis providing insight into reusing state-of-the-art stereo-matching algorithms with wide-angle rectified images. First, we tackle the error propagation using a model which better represents omnidirectional paradigm. Second, we conduct a comprehensive analysis of the effects of local distortions on the matching ability, and consequent distance estimates. The contribution of distortions are impossible to measure on existing datasets because it is impossible to separate the effects of scene variation and localized distortions. For that reason, we generate a unique dataset where the same scene is propagated throughout the whole image plane, providing scene-invariant localized analysis directly measuring the contribution of the distortion.
We evaluate 6 popular stereo-matching algorithms, from traditional to the latest state-of-the-art. Our work presents a necessary baseline for any researchers interested in obtaining omnidirectional distance estimations while benefiting from reusing state-of-the-art stereo-matching algorithms.

2. Related Work

2.1. Rectified stereo disparity estimation

The disparity estimation for rectified stereo is a long-standing problem in computer vision. The objective is to find a match for every pixel in the left image by searching along the same row in the right image. Traditionally, disparity estimation would be formulated as an optimization problem with several stages, such as cost computation, aggregation, refinement of the disparity [11] [37] and post-processing [27]. The fundamental limitation of the early methods was the handcrafted nature of the algorithms, suffering from false matches and the inability to deal with low-textured areas[39]. The earliest end-to-end deep learning approach by Mayer [19] has been based on the Flownet architecture for optical flow [4]. One of the most impactful ideas was the work of Lipson [18], who introduced an iterative refinement of the disparity maps simultaneously at multiple scales, also based on prior work on optical flow [30]. Since then, researchers have continued building on the iterative approaches with works such as DLNR [38] and IGEV-Stereo [36]. The most recent advancements is the utilization of the transformer models, such STTR [16] and Croco-Stereo [33].

2.2. Wide angle rectified stereo research

Abraham [1] initially developed the concept of fisheye stereo rectification, which involved calibrating a stereo pair and creating a virtual pair that follows epipolar constraints with equidistant sampling. They noted distortion near the poles in comparison to the raw image and suspected it could cause problems but lacked concrete evidence. Ohashi [24] conducted a study comparing pinhole and equirectangular projections to determine which was more suitable. Krombach [15] also explored equirectangular rectification and assessed different stereo-matching algorithms, but only traditional methods. Blaser [3] used Abraham's rectification approach for 360° urban scene reconstruction. The accuracy assessment was limited to comparisons with real objects of known geometry without measuring accuracy across different image areas. Schneider [29] conducted a

thorough evaluation of wide-angle stereo and its compatibility with standard stereo-matching algorithms, emphasizing the challenges in fisheye lens modelling. Beekmans [2] employed fisheye cameras for cloud distance measurement using equidistant projection. They analyzed distance variance but only in relation to a plane, a method we deem inadequate for omnidirectional distance estimation. Wang [32] has utilized two vertically omnidirectional cameras and designed a custom stereo-matching algorithm. Gao [6] has used two 245° cameras mounted opposite each other with a 65° overlap within which they used pinhole rectification. The most recent work by Xie [35] has deconstructed four wide-angle cameras into eight virtual pinhole images forming four rectified stereo pairs. While wide-angle stereo rectification is a recognized method, there currently does not exist any comprehensive analysis determining compatibility with the latest state-of-the-art stereo-matching algorithms.

2.3. Wide angle unrectified stereo research

The constraints on the FOV of the pinhole camera projection have led researchers to design rectification-less stereo-matching algorithms aiming to preserve the sensor-level image quality. However, this comes at the cost of searching for matches along epipolar lines, which no longer project to straight lines or utilising plane/sphere sweeping methods. Moreau [23] reconstructs a sparse pointcloud of the environment by utilising traditional feature matching on raw images by searching along curved epipolar lines. Hane [9] has reformulated the plane sweeping implementation, which works directly on non-rectified fisheye images. This approach can produce semi-dense depth maps. Roxas [25] has adopted the variational approach to work on raw fisheye images by searching along epipolar lines created by generating a trajectory field. Won [34] utilised a sphere-sphere-sweeping algorithm for a wide-baseline stereo mounted on a vehicle. Meuleman [22] has created a custom implementation of sphere-sweeping stereo with a fast inter-scale bilateral cost volume filtering, which improves performance in textureless regions. Kang [12] has implemented a stereo-matching paradigm for omnidirectional cameras by subdividing the camera space into a spherical geodesic grid. A traditional block-matching procedure was then implemented to work directly on the grid.

3. The projective geometry of wide-angle stereo-rectified images

stereo-matching algorithms provide disparity measurements relative to the reference frame of the stereo-rectified image plane. To align the epipolar lines with the rows of the image, the pixels need to have associated rays which are co-planar with the epipolar plane. The choice of angular sampling between the rays of pixels within the row and the angular sampling between the epipolar planes determines the relationship of disparity to the distance estimates. In perspective sampling, the inherent linearity guarantees that a given disparity yields a consistent depth estimate across the entire rectified image plane, though this comes at the expense of a reduced FOV. To address the limitation, we will utilize equidistant sampling instead as in [1]. As a consequence, the relationship between the estimated disparities and distances will be non-linear and while producing visual distortions, both of which are challenging the assumptions made in designing common dense stereo methods. Therefore, our analysis will focus on these two elements. Firstly, we introduce necessary equations in section 3.1 and 3.2. Section 3.4 examines the propagation of errors from the disparity estimates to distance estimates. Afterwards, a unique data-driven experiment will be used to measure the precise impact of the distortion on the accuracy of disparity estimation, as well as distance estimation using six state-of-the-art stereo-matching algorithms in section 4.

3.1. Projective geometry of wide-angle cameras

A pre-requisite for manipulating stereo image is accurate calibration of the wide-angle stereo pair. A camera model with correct intrinsic parameters establishes the relationship between each pixel and corresponding ray direction, which will be used to generate the equidistant stereo-rectified representation. There are numerous calibration toolboxes [26][31][17][20][10][5] that can be used to recover accurate parameters for wide-angle cameras. We will use Kannala-Brandt model [13], but the choice does not affect the outcomes, as long as it can accurately represent given wide-angle lens. The following subsection will describe how a 3D ray relative to the camera's optical frame ${}^{\text{Optical frame}}\text{Ray}$ can be mapped to the image coordinates ${}^{\text{Image frame}}\text{P}$, illustrated in figure 2.

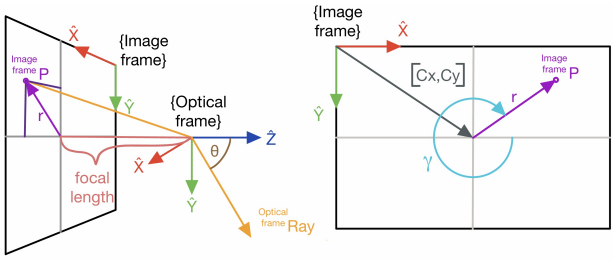

Figure 2. The projective geometry of the equidistant camera model.

Let ${}^{\text{Optical frame}}\text{Ray} = [x, y, z]^T$ and $r = \sqrt{x^2 + y^2}$. Then the projection function $\boldsymbol{\pi}\left[{}^{\text{Optical frame}}\text{Ray}, \mathbf{i}\right]$ maps a ray vector ${}^{\text{Optical frame}}\text{Ray}$ to a point in the image frame ${}^{Imageframe}\text{P}$ using intrinsic parameters $\mathbf{i} = \begin{bmatrix} fx & fy & Cx & Cy & k1 & k2 & k3 & k4 \end{bmatrix}^{\mathsf{T}}$

$$\boldsymbol{\pi}\left[{}^{\text{Optical frame}}\text{Ray}, \mathbf{i}\right] = \begin{bmatrix} f_x \theta_d \frac{x}{r} \\ f_y \theta_d \frac{y}{r} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \quad (1)$$

Where the distortion θ_d applied to the angle $\theta = atan2(r, z)$ is modeled as:

$$\theta_d = \theta(1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + k_4\theta^8) \quad (2)$$

3.2. Epipolar rectification

The concept of a virtual camera is used to generate an ideal stereo pair, where both cameras are only displaced along the ${}^{\text{Rectified frame left}}\hat{x}$ directions and all the raxels [8] within a row are complanar as illustrated in 3. Then each raxel can be mapped into sub-pixel location in the raw image using equations 3.1, generating a lookup table which can be reused for every new captured image. Real stereo pairs will typically lack ideal properties, and therefore, each raxel has to be transformed to account for incorrect rotation using extrinsic parameters extracted during calibration.

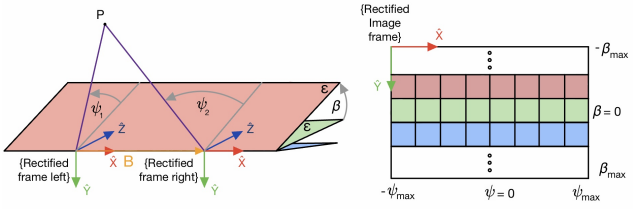

Figure 3. Epipolar constraints geometry of a stereo pair.

Let us assume that the stereo-rectified images are of a shape $[dim_x, dim_y]$ with the principal point in the centre $[C_x, C_y] = [dim_x/2, dim_y/2]$. Then, we can assume that the angle within the epipolar plane ψ = f(x*) where x* = x - C_x and epipolar angle β = f(y*) where y* = y - C_y, and y*, x* are representing the conditioned coordinates expressed relative to the centre of projection.

For equidistant representation, the angular sampling between the raxels is constant. For example, if the FOV within the epipolar plane is 180°, meaning ψ_{max} = 90°, the rectified stereo image focal length $f_{\text{x}} = C_x/\psi_{\text{max}}$. Similarly, β_{max} = 90°, $f_{\text{y}} = C_y/\beta_{\text{max}}$. Consequently, a coordinate at a point x,y in the rectified image plane has an associated ray vector ${}^{\text{Rectified frame}}\text{Ray}$, generated by equations 3 and 4.

$$\psi = x^*/f_{\text{x}}, \beta = y^*/f_{\text{y}} \quad (3)$$

$${}^{\text{Rectified frame}}\text{Ray} = \begin{bmatrix} \cos(\psi) \\ -\sin(\psi)\sin(\beta) \\ \sin(\psi)\cos(\beta) \end{bmatrix} \quad (4)$$

Conventional stereo-matching algorithms can be directly employed on the equidistant rectified representation, generating a disparity D for each pixel [x,y]. Then, the direction vectors ψ_1 and ψ_2 can be recovered using equation 3 for coordinates [x,y] in the left image and [x-D,y] in the right image. The 3D coordinates corresponding to a pixel [x,y] can be calculated using equations 5 and 6, where B stands for baseline.

$${}^{\text{Epipolar plane}}\text{P} = \left(\frac{B\sin(\psi_1)\cos(\psi_2)}{\sin(\psi_1 - \psi_2)}, \frac{B\cos(\psi_1)\cos(\psi_2)}{\sin(\psi_1 - \psi_2)}\right) \quad (5)$$

$${}^{\text{Rectified frame}}\text{P} = \begin{bmatrix} {}^{\text{Epipolar plane}}\text{P}_{\text{x}} \\ {}^{\text{Epipolar plane}}\text{P}_{\text{y}}\sin(\beta) \\ {}^{\text{Epipolar plane}}\text{P}_{\text{y}}\cos(\beta) \end{bmatrix} \quad (6)$$

3.3. Understanding distortion induced by stereo rectification

While the intent is to utilise the stereo-rectified images for processing, it is essential to examine how is the raw image sampled to produce them. Figure 5 shows the back-projection of the raxels, selecting rows corresponding to 15° increments of epipolar angles β. Furthermore, a jet colour scheme represents the angles ψ within the epipolar plane.

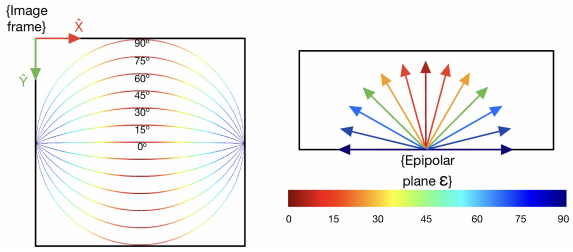

Figure 5. Visualisation of epipolar lines on raw source equidistant image.

Enforcing epipolar constraints results in a non-homogeneous sampling of the raw image as the ψ approach 90°, which is a singularity. While commonly referred to as distortions, the visual effects around the poles of the rectified images are a consequence of increasingly denser sampling

3.4. Distance uncertainty estimation for wide-angle rectified stereo

The performance of stereo-matching algorithms developed for perspective images is assessed based on their correspondence-finding ability in the stereo-rectified image plane. The most common metric counts the number of disparity estimates worse than a threshold (1,2,3) px. For rectified stereo vision, it makes sense to concentrate on disparity metrics because there is a direct, linear relationship between disparity and depth. This means that any specific disparity value observed in the stereo images will correspond to the

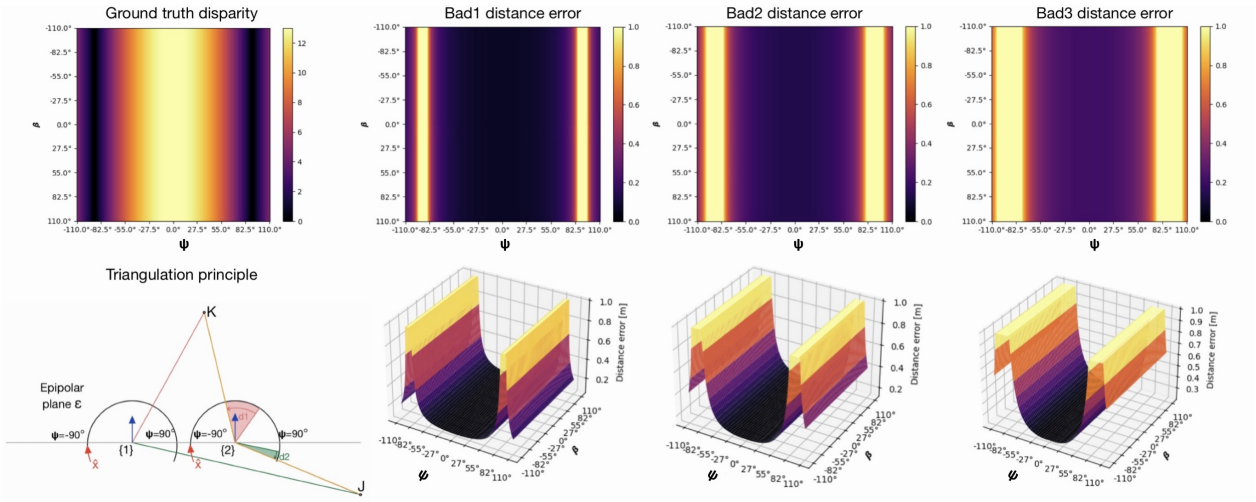

Figure 4. Top left figure visualizes disparity magnitude based on the spherical distance setting described in section 3.4. The illustration in the bottom left demonstrates the change of sign of disparity as ψ goes beyond $\pm 90^\circ$. The last three columns visualize the expected distance error given that a ground truth disparity is incorrect by (1,2,3) px for ψ and β. The units of error for the last three columns are in meters.

same depth measurement, regardless of its position in the image.

When utilizing an equidistant stereo-rectification model, the relationship between computed depth and disparity estimates becomes non-linear and changes throughout the image plane. The non-linear relationship has been a studied in prior art. Beekmans [2] evaluated the depth estimation accuracy of an equidistant rectified stereo pair relative to a plane. In our analysis, we will consider analysis relative to a surface of a sphere, as that represents the omnidirectional distance estimation paradigm more accurately. Schneider [29] has evaluated the accuracy of distance estimates relative to the incident angle with the optical axis. We evaluate the accuracy relative to ψ and β, considering the apparent bilateral symmetry of the equidistant rectified images.

Figure 4 outlines the expected ground truth disparity for a wide-angle stereo setup spanning $\pm 110^\circ$ in ψ and β relative to a surface of a sphere with a radius of 3 m. The effects of the singularity can be observed as the ψ approaches $\pm 90^\circ$, where the disparity becomes 0. The result is intuitive as the effective baseline also becomes 0, but also points to the necessity of separating the analysis relative to ψ and β rather than radially symmetric θ as in [29].
In the theme of reusing the existing rectified stereo disparity estimation algorithms, we introduce disparity error in the rectified image plane and observe the resulting distance error. The results can be seen in the last three columns in figure 4. The distance error is relatively flat in small values of $\psi = \pm 60^\circ$, increasing exponentially around $\psi \pm 90^\circ$, which have been clipped when exceeding the error of 1 m. Furthermore, when ψ exceeds $\pm 90^\circ$, the disparity sign flips making it incompatible with existing disparity estimation algorithms as they search only in one direction. The principle can be seen in the bottom left of figure 4.

The analysis of distance measurement accuracy was purely geometry-based, disregarding the dynamics between rectified stereo-matching algorithms and non-homogenous sampling induced distortions outlined in section 3.3 Regardless, it provides valuable insights into how to carry out an analysis that includes measuring the effects of distortions, which will be carried out in the section 4.

4. Data-Driven Evaluation

Exclusively geometric analysis from section 3.4 addressed the expected distance errors given homogeneous disparity error. While the problem formulation is the same for perspective and wide-angle stereo images; finding correspondences along the rows of the images, wide-angle stereo images present an additional layer of image distortion on top of already distorted fish-eye images. The change of visual appearance of the stereo-rectified images presents deviation from the domain for which perspective stereo-matching algorithms have been designed, and therefore has unknown consequences.
Measuring only the effects of distortion on the accuracy of disparity estimates is challenging. The source of errors in disparity estimates is primarily the difficulty of the scene. For example, a complicated object such as a chair with many disparity discontinuities and thin structures might have more significant disparity errors in the centre of the image than a well-textured flat surface in a more distorted area. For this reason, none of the available wide-angle stereo datasets are suitable for such analysis.

To address this shortcoming, we designed a unique ex-

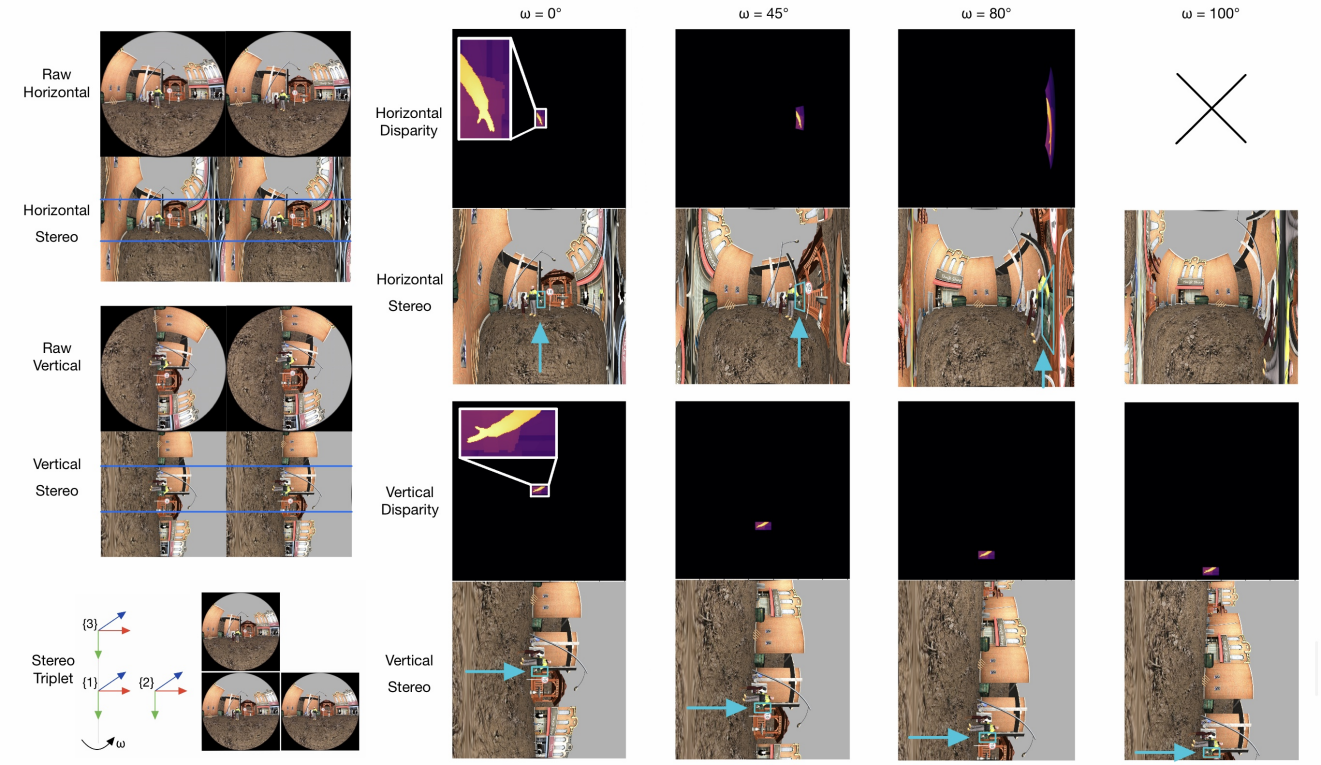

Figure 6. The core principle of our data acquisition pipeline for isolating the measurements of the distortion. As we rotate the stereo triplet by ω, a challenging scene segment will be represented throughout the image plane in both ψ and β directions.

periment that allows for measuring only the error contribution of distortion while maintaining a challenging scene with small objects, strong disparity discontinuities, and occlusions. We build a challenging scene in Gazebo [14] simulation with high resolution models from the ignition database. Then, we select a challenging segment of 10° by 20° and render ground truth point-cloud of it. The scene and the rendered ground truth segment can be seen in figure 7.

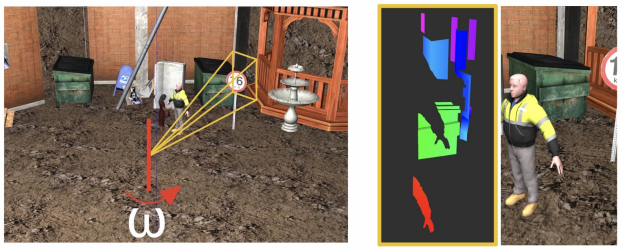

Figure 7. The ground truth acquisition pipeline.

The key to our evaluation methodology is rotating a stereo camera by 1° increments. As a consequence of the rotation ω, the scene segment will be represented throughout the image plane, from the middle to the very edge. To build on purely geometrical analysis from the section 3.4, it is essential to explore the bilateral symmetry of the rectified stereo image plane. For that reason, we set up a stereo-triplet, which can be seen in the bottom left of figure 6. The stereo camera follows the equidistant projection model and has a 1440 x 1440 px resolution. Raw images have FOV of 220° and rectified images span $\pm 110°$ in ψ and β. The configuration forms two orthogonal stereo pairs. As the whole triplet is rotated around ω, the scene segment will transition over the range of ψ for the horizontal stereo pair formed by cameras (1,2). Similarly, the scene segment will transition throughout the range of β for the vertical stereo pair formed by cameras (1,3). A transition over ψ contains a rapid change as it approaches the singularity at 90°. For this reason, the scene segment is more narrow in the ψ direction to better capture the nuance.

Figure 6 further showcases examples of ground truth disparity for various values of ω. The horizontal stereo exhibits strong stretching of the narrow segment as it transitions over the range of ψ as a consequence of rotation around ω, especially noticeable close to the singularity. We chose to omit generating ground truth disparity values for angles of ψ of more than $\pm 90°$ as the resulting disparity would change sign. Furthermore, the areas beyond ψ of more than $\pm 90°$

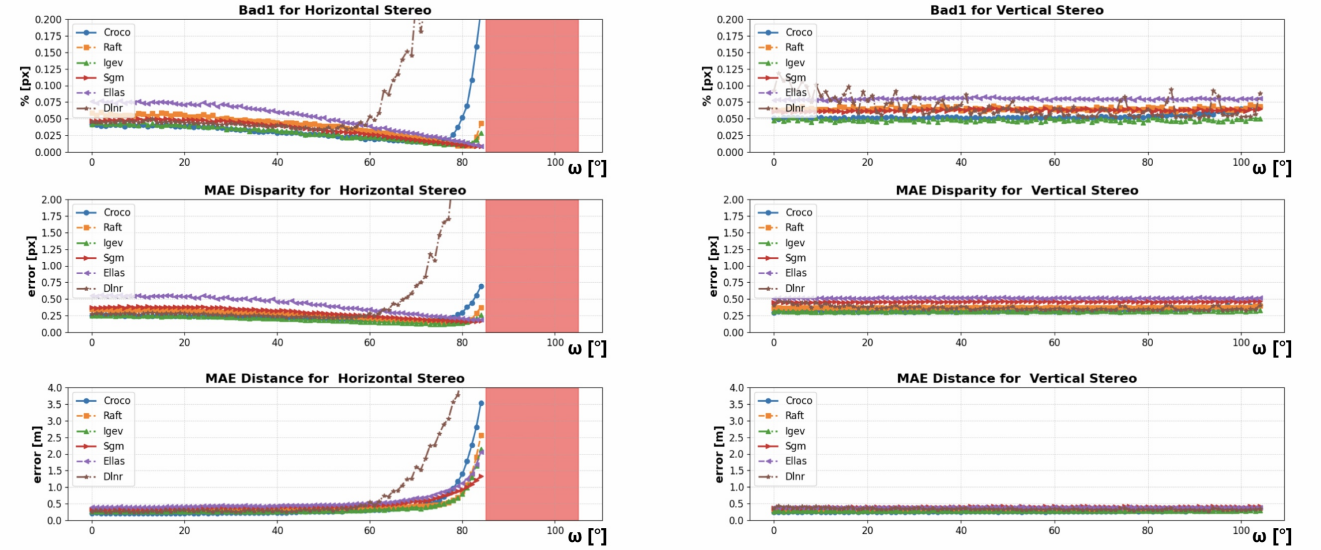

Figure 8. The evaluation of the accuracy of disparity and distance estimation as a challenging scene transitions throughout the rectified stereo image plane. Every metric has been computed for each scene segment individually and then averaged to show a clearer signal.

would not be observable due to occlusion of the other camera from the stereo pair in real applications.

On the other hand, the vertical stereo ,where the segment transitions through β, shows no sign of distortion within the whole range. Moreover, the whole range maintains the same sign within disparity values, making it compatible with existing rectified stereo algorithms. For that reason, we render the ground truth disparity up to β of $110°$. Such a setting is also compatible with real stereo cameras, as the occlusion would not be an issue. Given that the scene segment is $10°$ wide and centered around ω, we generate the dataset relative to ω and generate samples up to the maximum boundary - $5°$. As such, each scene segment is captured by 85 stereo images in horizontal stereo and 105 images in vertical stereo. A summary about the size of our dataset and experimentation can be seen in figure 1.

Table 1. Dataset Overview for Evaluation

Aspect	Horizontal	Vertical
Segments	36	36
Images per Segment	85	105
Total Evaluation Examples	3060	3780

4.1. Evaluation of the collected dataset

To answer the question of compatibility between wide-angle rectified images, we have selected 6 popular stereo-matching algorithms while representing different design methodologies. From traditional algorithms, we have chosen ELLAS [7] and Semi-global matching [11]. From deep learning based methods, we have chosen the original implementation of RAFT-Stereo [18] and its more advanced variations IGEV-Stereo [36] and DLNR [38]. Lastly, we include Croco-Stereo [33] as it is based on a novel transformed based architecture. All deep learning models were used with weights dedicated to the Middlebury dataset with associated inference parameters defined by authors in the original papers. Semi-global matching used a window size of 5, and the penalties P1 and P2 based on empirical observations and the characteristics of our dataset. The evaluation was performed for all computed disparities. All deep learning based methods were fully dense, while SGM and ELLAS omit some pixels by design.

Figure 8 summarizes the results of the evaluation aimed at addressing the contribution of the distortion. The results are averaged over all 36 non-overlapping scene segments from the same scene. To interpret the results, it is essential to consider only the trend over the range ω. Absolute measurements would change based on different baseline, rectified image resolution or varying the difficulty of the scene segments. A surprising observation is a decreasing trend of bad1 px error, as well as disparity MAE for the horizontal stereo as ω approaches $85°$ for most of the algorithms. Considering the ground truth disparity figure from 4, the decreasing values of the ground truth disparity might be making it easier to find correspondences. On the other hand, algorithms such as DLNR and Croco-Stereo exhibit low accuracy at angles beyond $60°$ and $75°$ respectively, demonstrating their inability to adapt to distortions. Regardless of the decreasing disparity error trend, the MAE of distance starts raising beyond $60°$ which is simply caused

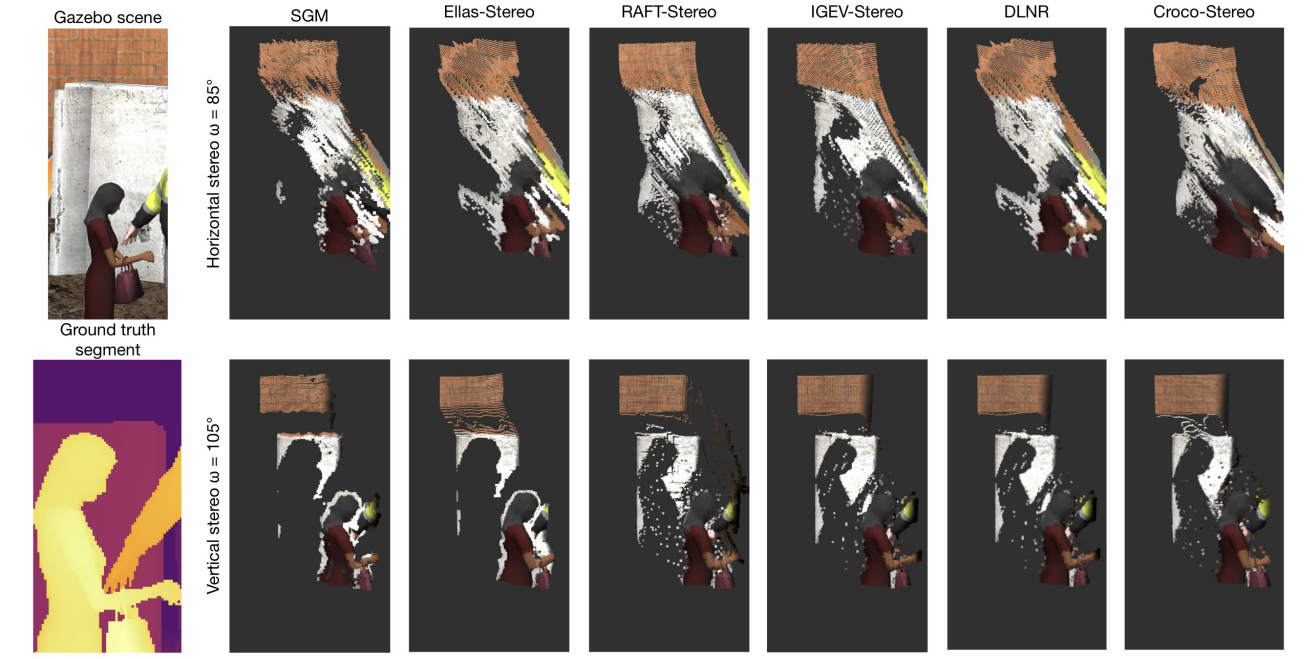

Figure 9. Visualising errors at extreme values of ω for vertical and horizontal stereo. A challenging scene segment has been chosen, demonstrating strong disparity discontinuities.

by disadvantageous non-linear triangulation outlined in section 3.4.

Regarding the vertical stereo, it can be observed that the bad1 px and MAE for disparity are homogeneous throughout the whole range. Such a result is expected, as the visuals show very little distortion as the scene segment transitions throughout the whole range of ω. Furthermore, as the triangulation geometry is advantageous throughout the whole range of β and small values of ψ, the distance error is also homogeneous.
Figure 9 outlines a qualitative evaluation of the computed pointcloud for one of the challenging scene segments. It can be observed that while the horizontal stereo fails to generate an accurate pointcloud, the vertical stereo manages to produce high accuracy results for every stereo-matching method, even at extreme angles of ω.

5. Discussion

The results of our experiments demonstrate that there are two fundamental limits to ultra-wide stereo perception - distortions and unfavourable triangulation geometry. These phenomena manifest in the same areas of the rectified image plane. Such a conclusion questions the motivation behind the research direction attempting to overcome the distortion caused by wide stereo-rectification and designing algorithms that operate directly on raw fisheye images. Furthermore, placing stereo cameras horizontally might not be the most effective option for wide-angle distance estimation, where most of the desired objects lie on a plane, a scenario common in autonomous driving and robotics.
Typically, horizontal stereo always has an occluded area on the very left side of the left image. By placing the cameras vertically, this area gets moved to an area where no objects of interest lie, such as the sky. Given that a stereo setup will always have bilaterally symmetric properties, utilizing a trifocal stereo setup in orthogonal configuration might offer a way to cancel out the disadvantageous triangulation properties and achieve high-accuracy disparity estimation throughout the whole FOV while being able to reuse existing algorithms.

6. Conclusion

We have conducted a comprehensive analysis of the interaction between state-of-the-art stereo matching algorithms and wide-angle stereo rectification. The results clearly show that omnidirectional distance estimation is possible without any need to adjust the correspondence matching algorithms. We have carried out a unique evaluation that has allowed us to isolate the effects of distortion while maintaining a challenging scene, demonstrating that the fundamental limitation is triangulation geometry. Our analysis provides actionable insights that can be leveraged to achieve high-accuracy omnidirectional distance estimation.

Funding This research was funded by Innovation Fund Denmark, grant number 3129-00060B.

References

[1] Steffen Abraham and Wolfgang Förstner. Fish-eye-stereo calibration and epipolar rectification. *Isprs Journal of Photogrammetry and Remote Sensing*, 59:278–288, 2005. 2, 3

[2] Christoph Beekmans, Johannes Schneider, Thomas Läbe, Martin Lennefer, C. Stachniss, and Clemens Simmer. Cloud photogrammetry with dense stereo for fisheye cameras. *Atmospheric Chemistry and Physics*, 16:14231–14248, 2016. 2, 3, 5

[3] Stefan Blaser, Stephan Nebiker, and Stefan Cavegn. System design, calibration and performance analysis of a novel 360° stereo panoramic mobile mapping system. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 207–213, 2017. 2

[4] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Häusser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2758–2766, 2015. 2

[5] Paul Timothy Furgale, Jörn Rehder, and Roland Y. Siegwart. Unified temporal and spatial calibration for multi-sensor systems. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286, 2013. 3

[6] Wenliang Gao and Shaojie Shen. Dual-fisheye omnidirectional stereo. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6715–6722, 2017. 3

[7] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Asian Conference on Computer Vision*, 2010. 1, 7

[8] Michael D. Grossberg and Shree K. Nayar. The raxel imaging model and ray-based calibration. *International Journal of Computer Vision*, 61:119–137, 2005. 4

[9] Christian Häne, Lionel Heng, Gim Hee Lee, Alexey Sizov, and Marc Pollefeys. Real-time direct dense matching on fisheye images using plane-sweeping stereo. *2014 2nd International Conference on 3D Vision*, 1:57–64, 2014. 2, 3

[10] Lionel Heng, Bo Li, and Marc Pollefeys. Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1793–1800, 2013. 3

[11] Heiko Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE transactions on pattern analysis and machine intelligence*, 30 2:328–41, 2008. 1, 2, 7

[12] Dong Hun Kang, Hyeonjoong Jang, Jungeon Lee, C. M. Kyung, and Min H. Kim. Uniform subdivision of omnidirectional camera space for efficient spherical stereo matching. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12962–12970, 2022. 2, 3

[13] Juho Kannala and Sami Sebastian Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1335–1340, 2006. 3

[14] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, pages 2149–2154 vol.3, 2004. 6

[15] Nicola Krombach, David Droeschel, and Sven Behnke. Evaluation of stereo algorithms for obstacle detection with fisheye lenses. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 33–40, 2015. 2

[16] Zhaoshuo Li, Xingtong Liu, Nathan Drenkow, Andy Ding, Francis X. Creighton, Russell H. Taylor, and Mathias Unberath. Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6197–6206, 2021. 2

[17] Derek D. Lichti, David Jarron, Wynand Tredoux, Mozhdeh Shahbazi, and Robert S. Radovanovic. Geometric modelling and calibration of a spherical camera imaging system. *The Photogrammetric Record*, 35, 2020. 3

[18] Lahav Lipson, Zachary Teed, and Jia Deng. Raft-stereo: Multilevel recurrent field transforms for stereo matching. *2021 International Conference on 3D Vision (3DV)*, pages 218–227, 2021. 2, 7

[19] Nikolaus Mayer, Eddy Ilg, Philip Häusser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4040–4048, 2015. 2

[20] Christopher Mei and Patrick Rives. Single view point omnidirectional camera calibration from planar grids. *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3945–3950, 2007. 3

[21] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1

[22] Andreas Meuleman, Hyeonjoong Jang, Daniel S. Jeon, and Min H. Kim. Real-time sphere sweeping stereo from multiview fisheye images. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11418–11427, 2021. 2, 3

[23] Julien Moreau, Sébastien Ambellouis, and Yassine Ruichek. Equisolid fisheye stereovision calibration and point cloud computation. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 167–172, 2013. 2, 3

[24] Akira Ohashi, Fumito Yamano, Gakuto Masuyama, Kazunori Umeda, Daisuke Fukuda, Kota Irie, Shuzo Kaneko, Junya Murayama, and Yoshitaka Uchida. Stereo rectification for equirectangular images. *2017 IEEE/SICE International Symposium on System Integration (SII)*, pages 535–540, 2017. 2

[25] Menandro Roxas and Takeshi Oishi. Real-time variational fisheye stereo without rectification and undistortion. *ArXiv*, abs/1909.07545, 2019. 2, 3

[26] Davide Scaramuzza, Agostino Martinelli, and Roland Y. Siegwart. A flexible technique for accurate omnidirec-

tional camera calibration and structure from motion. *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, pages 45–45, 2006. 3

[27] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2001. 2

[28] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nesic, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, 2014. 1

[29] Johannes Schneider, C. Stachniss, and Wolfgang Förstner. On the accuracy of dense fisheye stereo. *IEEE Robotics and Automation Letters*, 1:227–234, 2016. 2, 5

[30] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision*, 2020. 1, 2

[31] Steffen Urban, Jens Leitloff, and Stefan Hinz. Improved wide-angle, fisheye and omnidirectional camera calibration. *Isprs Journal of Photogrammetry and Remote Sensing*, 108: 72–79, 2015. 3

[32] Ning-Hsu Wang, Bolivar Solarte, Yi-Hsuan Tsai, Wei-Chen Chiu, and Min Sun. 360sd-net: 360° stereo depth estimation with learnable cost volume. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 582–588, 2019. 3

[33] Philippe Weinzaepfel, Thomas Lucas, Vincent Leroy, Yohann Cabon, Vaibhav Arora, Romain Brégier, Gabriela Csurka, Leonid Antsfeld, Boris Chidlovskii, and Jérôme Revaud. CroCo v2: Improved Cross-view Completion Pre-training for Stereo Matching and Optical Flow. In *ICCV*, 2023. 2, 7

[34] Changhee Won, Jongbin Ryu, and Jongwoo Lim. Sweepnet: Wide-baseline omnidirectional depth estimation. *2019 International Conference on Robotics and Automation (ICRA)*, pages 6073–6079, 2019. 3

[35] Sheng Xie, Daochuan Wang, and Yunhui Liu. Omnividar: Omnidirectional depth estimation from multi-fisheye images. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21529–21538, 2023. 3

[36] Gangwei Xu, Xianqi Wang, Xiao-Hua Ding, and Xin Yang. Iterative geometry encoding volume for stereo matching. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21919–21928, 2023. 2, 7

[37] Li Zhang and Steven M. Seitz. Estimating optimal parameters for mrf stereo from a single image pair. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29: 331–342, 2007. 2

[38] Haoliang Zhao, Huizhou Zhou, Yongjun Zhang, Jing Chen, Yitong Yang, and Yong Zhao. High-frequency stereo matching network. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1327–1336, 2023. 2, 7

[39] Kun Zhou, Xiangxi Meng, and Bo Cheng. Review of stereo matching algorithms based on deep learning. *Computational Intelligence and Neuroscience*, 2020, 2020. 2