Supplementary Material: All-directional Disparity Estimation for Real-world QPD Images

Hongtao Yu[1] Shaohui Song[1] Lihu Sun[1] Wenkai Su[1] Xiaodong Yang[2] Chengming Liu[2]
[1] OMNIVISION, Wuhan, China [2] OMNIVISION, Santa Clara, CA

In this supplementary material, firstly, we examine the asymmetry of the PSF in real QPD sensors (Sec. 1). Secondly, we provide a comprehensive overview of the QPD2K dataset (Sec. 2), encompassing QPD2K samples, disparity ground truth acquisition and a summary of QPD datasets. Thirdly, we provide the reparameterization method for the illumination-invariant module and models' complexity (Sec. 3) and finally present additional qualitative and quantitative results on the QPD2K dataset (Sec. 4).

1. PSF Calibration

Theoretically, the point spread function (PSF) of a dual-pixel (DP) sensor is symmetric under ideal conditions [6, 11]. However, the PSF of actual DP sensors deviates from this idealized model, exhibiting spatial variance and asymmetry [14].

To examine the symmetry of the PSF in real QPD sensors, we utilize the QPD sensor (OV50A) to measure the PSF by a well-established calibration method [1, 8]. The calibration pattern comprises a grid of small disks and blocks. Given that our camera features a fixed optical system, we obtain various in-focus and out-of-focus image pairs by adjusting the position of the voice coil motors (VCMs). Using the in-focus and out-of-focus image pairs, we compute the absolute PSFs according to Eq. (1). We restrict our calculation to the central region to minimize the distortion effect in the border region of images. The calibration pattern and PSFs are illustrated in Fig. 1.

$$\arg\min_{h} \sum_{j=1}^{n} \lambda_j \left\| f_j * (i_s * h - i_B) \right\|_2^2 + \lambda_{n+1} \left\| \bigtriangledown h \right\|_2^2 + \lambda_{n+2} \left\| R \circ h \right\|_2^2, \tag{1}$$

where i_B and i_S denote the out-of-focus and in-focus images respectively, R is spatial regularization matrix, The constraints $\lambda_j, \lambda_{n+1}, \lambda_{n+2}$ ensure that the kernel is non-negative. The estimated PSF kernel is represented by h, and f_j is a filter applied to the images. Our calibration results indicate that the PSFs are asymmetric in real QPD sensors.

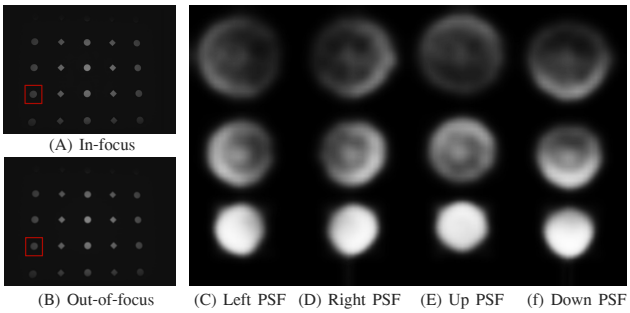


Figure 1. (A) In-focus calibration pattern (B) Out-of-focus calibration pattern. (C) Different blur degrees from top to bottom. Notably, our PSF calibration results also reveal spatial variance, non-circularity and asymmetry.

2. More Datasets Details

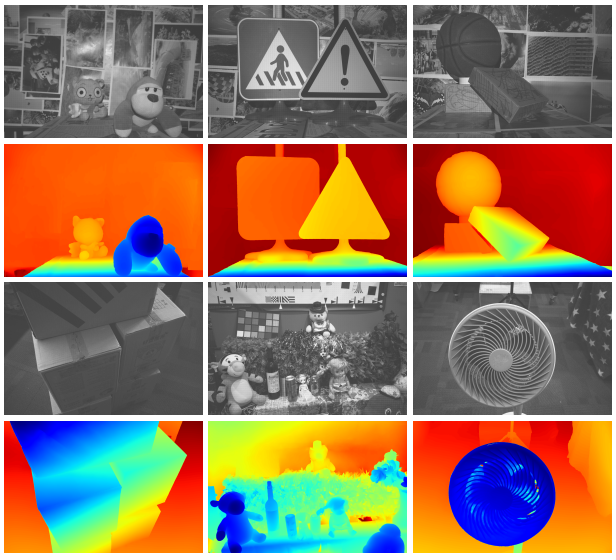


Figure 2. QPD2K samples: QPD raw data and the corresponding disparity ground truth.

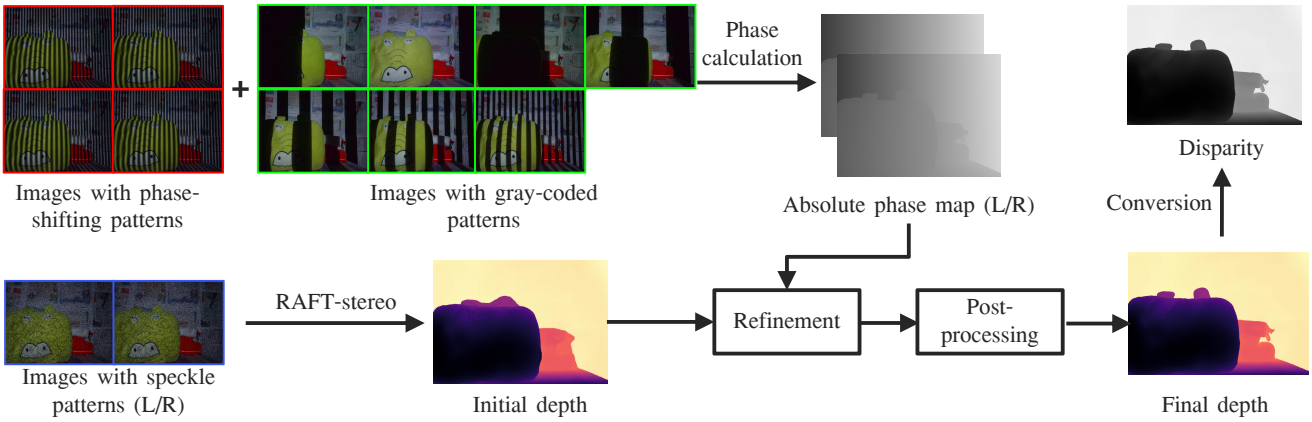


Figure 3. Disparity ground truth acquisition pipeline.

2.1. QPD2K Samples

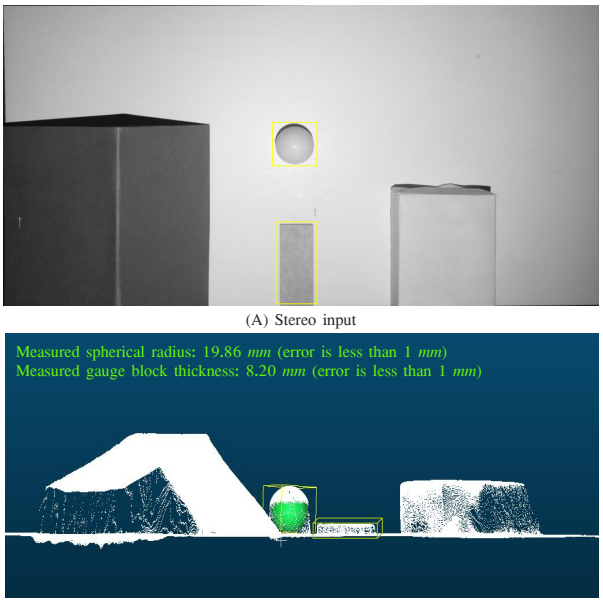
In the main text, we described the configuration of the dataset. We provide additional samples from the training set of our QPD2K dataset, illustrated in Fig. 2. Notably, our dataset comprises complex scenes with cluttered backgrounds and expansive natural environments.

2.2. Disparity Ground Truth Acquisition.

Directly estimating the disparity of a natural scene is a challenging task, whereas acquiring the depth information is relatively more straightforward. To overcome this limitation, we first acquire depth information and subsequently calibrate the disparity-depth relationship of QPD data to acquire disparity information. In the following section, we will elaborate on our pipeline, disparity calculation details, and measurement accuracy to provide a more comprehensive understanding of our dataset.

Overall pipeline. The workflow of our algorithm is depicted in Fig. 3. To achieve high-precision disparity, we employ a multi-stage approach. Initially, we project speckle structured light to obtain a coarse depth map. However, this method is susceptible to inaccuracies in ill-posed regions. To address this limitation, we leverage the texture-invariant properties of structured light and project a sequence of seven Gray-coded patterns and four phase-shifting sinusoidal patterns to generate an absolute phase map. We then refine the depth results by matching the phase map within the range of the initial depth, yielding an optimized depth map. We further refine the depth map through outlier removal and median filtering. Ultimately, we acquire disparity information from refined depth by the disparity-depth relationship.

Depth accuracy. To assess the measurement accuracy of the QPD disparity acquisition system, we perform an experiment involving the measurement of standard spheres and a gauge block. As illustrated in Fig. 4, the actual radius of the standard spheres is $20\,mm$, whereas our measured radius is $19.86\,mm$, resulting in a measurement error of $0.14\,mm$. Similarly, the actual height of the gauge block is $9\,mm$, and our measured height is $8.2\,mm$, yielding a measurement error of $0.8\,mm$. These results demonstrate the high precision of our system.



(A) Stereo input

(B) Depth (point cloud visualization)

Figure 4. By measuring standard spheres and gauge block, our results show a mean error of less than 1 mm, which validates the high precision of our method.

Disparity calculation details. The core of QPD disparity ground truth calculation involves establishing correspondences between the left and right or up and down sub-image pairs. To achieve this, we employ the Normalized Cross-Correlation (NCC) method to compute disparity, which is robust to illumination variations. The QPD dataset exhibits

Datasets	Input type	Ground-truth type	Ground-truth acquisition	Scenarios	Resolution
DP-disp [11]	Dual-pixel	Depth	Depth-from-defocus	Indoor/Outdoor (only testset)	5180×2940
DP-Face [5]	Dual-pixel	Depth	Structured-light	Human face	1680×1120
DP-PixelPhone [4]	Dual-pixel	Depth	Multi-view stereo	Indoor/Outdoor	1512×2016
QP-Data [13]	Quad-pixel	Disparity	Simulation	Indoor/Outdoor	1024×768
DP5K [7]	Dual-pixel	Depth	Structured-light	Indoor	1024×768
dpMV [3]	Dual-pixel	Depth	Multi-view stereo	Indoor/Outdoor	1512×2016
QPD2K (Ours)	**Quad-pixel**	**Disparity**	**Stereo structured-light**	Indoor	3000×2000

Table 1. Existing DP/QPD datasets summary. Our QPD2K dataset provides high-resolution, real-world disparity ground truth.

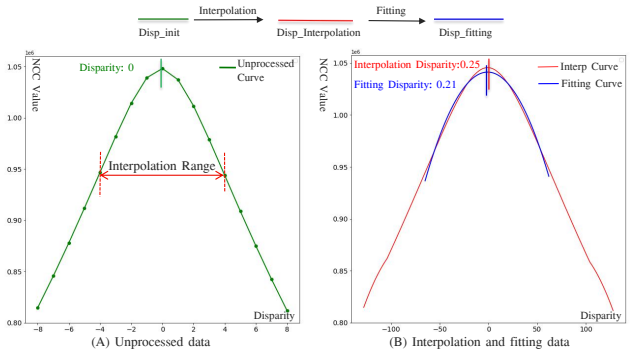


Figure 5. Disparity calculation pipeline.

dual-directional information, and we consider the horizontal disparity as an example. The pipeline for calculating disparity is illustrated in Fig. 5. Firstly, we divide the left view into non-overlapping 64×64 tiles and calculate the disparity for each tile by setting the search range to $[-8, 8]$ with a step size of 1, resulting in 17 points of NCC matching costs. We identify the maximum matching cost correspondence disparity D_{init}. To obtain the sub-pixel disparity, we interpolate the NCC matching cost using a Bezier curve interpolation around D_{init} within the range of $[-8, 8]$, increasing the number of points from 17 to 257. We then determine the maximum matching cost correspondence disparity D_{sub}. Ideally, the NCC matching cost curve should follow a parabolic distribution, and we perform a quadratic fit to find the maximum value and corresponding disparity D_h. We compute the confidence level $Conf_h$ by comparing the fitted curve with the original NCC matching cost. Repeating these steps for the vertical disparity, we obtain the disparity D_v and corresponding confidence $Conf_v$. Finally, we combine the two directions' disparity using confidence-weighted fusion to attain the final disparity D_f:

$$D_f = \frac{D_h \times Conf_h + D_v \times Conf_v}{Conf_h + Conf_v}. \tag{2}$$

This approach enables more accurate disparity estimation by leveraging the dual-directional information inherent in the QPD dataset.

Disparity-depth fitting. To mitigate the impact of lens distortion on the disparity-depth relationship, we divide the calibration image into 5×5 blocks and select the central 3×3 blocks to analyze the relationship between depth and disparity. Our fitting results are presented in Fig. 6. Notably, we observe that the disparity-depth relationship exhibits spatial variability, prompting us to restrict our analysis to the central region of the image with a resolution of 3000×2000. Furthermore, we apply a spatial averaging technique by taking the mean of 3×3 windows to establish a robust and accurate disparity-depth relationship.

2.3. QPD Dataset Summary.

We present a comprehensive summary of publicly available datasets related to DP and QPD sensors in Tab. 1. Notably, only QP-Data [13] and QPD2K offer disparity, whereas our dataset stands out as the sole high-resolution, real-world disparity dataset. Our dataset bridges the gap between QPD-based tasks and DP-based ones.

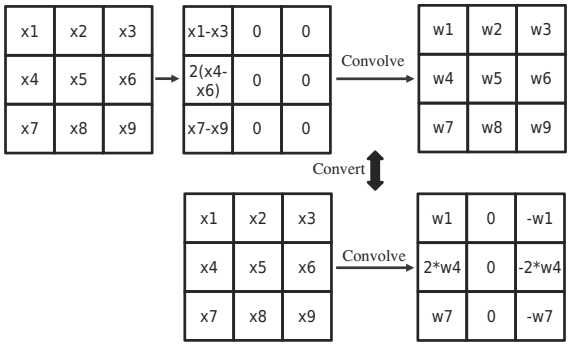
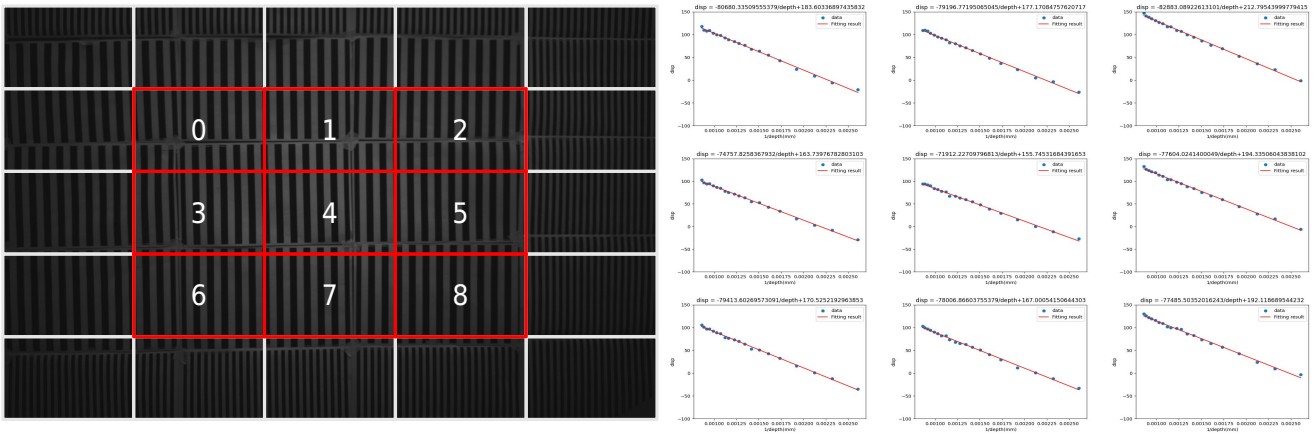


Figure 7. Convert HSDC into standard convolution.

3. Method Details

In this section, we provide additional details on the reparameterization method for the illumination-invariant module and analyze the model complexity.

Illumination-invariant module reparameterization. As depicted in Fig. 7, we exemplify our approach using the horizontal sobel differential convolution (HSDC), where we leverage the sobel operator to extract the differential features from the input. Subsequently, we apply a convolu-



(A) Calibration pattern

(B) Disparity-depth fitting

Figure 6. Disparity-depth fitting. The QPD disparity and the depth are in an affine relation.

tional operation to these differential features, resulting in the generation of the final differential features. After training, we compute the derivatives of the convolutional kernels, enabling the calculation of the differential features.

Model complexity analysis. In comparison to conventional DP-based disparity estimation methods, such as SFDB [6], and the QPD disparity estimation method, QPDNet [13], we present a comprehensive complexity analysis of our proposed model in Tab. 2. Additionally, the complexity analysis of the ablation study is summarized in Tab. 3, providing insights into the impact of individual components on the overall model complexity.

Method	Input type	Params (M)	Flops (G)
SFDB [6]	DP	10.64	1048
QPDNet [13]	QPD	13.69	1951
DPNet (Ours)	DP	18.43	1304
QuadNet (Ours)	QPD	24.43	2793

Table 2. Model complexity summary.

Method	Input type	Params (M)	Flops (G)
Base	DP	9.37	1153
Base + IIM	DP	12.54	1259
Base + Subpixel	DP	18.01	1218
Base + IIM/Subpixel	DP	18.43	1304
Base + All	QPD	24.43	2793

Table 3. Ablation experiments complexity summary.

4. Additional Results

We adopt the evaluation metrics employed in the Scene Flow dataset [9], but due to the relatively small disparity in our dataset, epe is as follows:

$$\frac{1}{N} \sum_{(x,y)\in N} |d_{est}\,(x,y) - d_{gt}\,(x,y)\,|. \tag{3}$$

Bad metrics are as follows:

$$\frac{1}{N_{all}} \sum_{(x,y)\in N_{all}} \{|d_{est}\,(x,y) - d_{gt}\,(x,y)\,| > \mathrm{P}\}, \tag{4}$$

where the P values of Bad0.3, Bad0.5, and Bad1 are 0.3, 0.5, and 1, respectively.

As shown in Fig. 8, we conduct additional comparisons between DPNet and QuadNet on QPD2K. In the main text, we have already demonstrated that our DPNet achieves the best performance among methods using DP input. However, since DP data can only provide horizontal disparity information, it often performs poorly in regions with weak textures or horizontally repetitive textures. In contrast, QuadNet, which can leverage both horizontal and vertical disparity information, achieves superior performance.

In Fig. 9, we further compare QuadNet with QPDNet [13]. For scenes with background, our QPDNet provides accurate edge and background disparity information, effectively handling overexposed and small-disparity regions where QPDNet often fails to estimate accurate background disparity. For open natural scenes, our QPDNet demonstrates a significant advantage in regions with weak textures and low light. Fundamentally, the difference lies in the fact that QPDNet performs fusion at the feature level, whereas our approach conducts fusion at the disparity level.

Based on the characteristics of the QPD2K testing set scenes, we categorize them into two types: scenes with background and open natural scenes. We present a quantitative comparison in Tab. 4. Our results indicate that QuadNet achieves superior performance in both categories. In contrast, QPDNet performs below expectations in scenes with the background. We attribute this to its estimation of only half of the disparity, which can lead to a degeneration in estimation accuracy.

(A) Reference (B) DPNet (Ours) (C) **QuadNet (Ours)** (D) GT

Figure 8. Additional qualitative experimental results on QPD2K. QuadNet demonstrates superior performance over DPNet in weak texture regions and repetitive structure.

(A) Reference (B) QPDNet [13] (C) **QuadNet (Ours)** (D) GT

Figure 9. Additional qualitative experimental results on QPD2K. Our QuadNet outperforms QPDNet [13] in ill-posed regions, such as over-exposure and blur areas.

Method	Scenes with background				Open natural scenes				Mean			
	bad 0.3↓	bad 0.5↓	bad 1↓	epe↓	bad 0.3↓	bad 0.5↓	bad 1↓	epe↓	bad 0.3↓	bad 0.5↓	bad 1↓	epe↓
DPdisp [11]	0.9900	0.9793	0.8045	1.5504	0.9363	0.8849	0.7382	1.8574	0.9691	0.9433	0.7791	1.6796
SFBD [6]	0.9765	0.9484	0.2188	1.0829	0.8515	0.7523	0.5431	1.2505	0.9266	0.8700	0.3665	1.1500
CCA [10]	0.9735	0.9528	0.4361	1.2776	0.8473	0.7505	0.5910	1.4797	0.9246	0.8743	0.4962	1.3560
QPDNet [13]	0.9675	0.9133	0.2114	0.9123	0.8208	0.7243	0.4809	1.0785	0.9088	0.8377	0.3372	0.9788
IEGV-Stereo [15]	0.9006	0.8359	0.6706	1.4239	0.9650	0.9374	0.8411	2.1620	0.9263	0.8765	0.7388	1.7192
Mc-stereo [2]	0.2682	0.1153	0.0263	0.1961	0.7927	0.6625	0.4131	1.3936	0.4780	0.3342	0.1810	0.6751
S-RAFT [12]	0.2837	0.1858	0.0099	0.2399	0.6685	0.4343	0.0791	0.5031	0.4376	0.2852	0.0376	0.3452
S-IEGV [12]	0.2637	0.1586	0.0221	0.2349	0.6981	0.4956	0.1899	0.6130	0.4375	0.2934	0.0892	0.3862
DLNR [16]	0.1005	0.0215	0.0031	0.1172	0.6718	0.4409	0.0790	0.5053	0.3290	0.1892	0.0335	0.2725
DPNet (Ours)	0.0489	0.0136	0.0029	0.0849	0.6095	0.3806	0.0789	0.4675	0.2731	0.1604	0.0333	0.2380
QuadNet (Ours)	**0.0280**	**0.0078**	**0.0016**	**0.0709**	**0.5683**	**0.2941**	**0.0396**	**0.4029**	**0.2294**	**0.1136**	**0.0166**	**0.2007**

Table 4. Additional quantitative results on QPD2K. Our QuadNet achieves superior performance in both categories.

References

[1] Abdullah Abuolaim, Mauricio Delbracio, Damien Kelly, Michael S Brown, and Peyman Milanfar. Learning to reduce defocus blur by realistically modeling dual-pixel data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 1

[2] Miaojie Feng, Junda Cheng, Hao Jia, Longliang Liu, Gangwei Xu, and Xin Yang. Mc-stereo: Multi-peak lookup and cascade search range for stereo matching. In *2024 International Conference on 3D Vision (3DV)*, pages 344–353. IEEE, 2024. 6

[3] Aryan Garg, Raghav Mallampali, Akshat Joshi, Shrisudhan Govindarajan, and Kaushik Mitra. Stereo-knowledge distillation from dpmv to dual pixels for light field video reconstruction. *arXiv preprint arXiv:2405.11823*, 2024. 3

[4] Rahul Garg, Neal Wadhwa, Sameer Ansari, and Jonathan T. Barron. Learning single camera depth estimation using dual-pixels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 3

[5] Minjun Kang, Jaesung Choe, Hyowon Ha, Hae-Gon Jeon, Sunghoon Im, In So Kweon, and Kuk-Jin Yoon. Facial depth and normal estimation using single dual-pixel camera. In *European Conference on Computer Vision*, pages 181–200. Springer, 2022. 3

[6] Donggun Kim, Hyeonjoong Jang, Inchul Kim, and Min H Kim. Spatio-focal bidirectional disparity estimation from a dual-pixel image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5023–5032, 2023. 1, 4, 6

[7] Feiran Li, Heng Guo, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Learning to synthesize photorealistic dual-pixel images from rgbd frames. In *2023 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2023. 3

[8] Fahim Mannan and Michael S Langer. Blur calibration for depth from defocus. In *2016 13th Conference on Computer and Robot Vision (CRV)*, pages 281–288. IEEE, 2016. 1

[9] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3061–3070, 2015. 4

[10] Sagi Monin, Sagi Katz, and Georgios Evangelidis. Continuous cost aggregation for dual-pixel disparity extraction. In *2024 International Conference on 3D Vision (3DV)*, pages 675–684. IEEE, 2024. 6

[11] Abhijith Punnappurath, Abdullah Abuolaim, Mahmoud Afifi, and Michael S Brown. Modeling defocus-disparity in dual-pixel sensors. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2020. 1, 3, 6

[12] Xianqi Wang, Gangwei Xu, Hao Jia, and Xin Yang. Selective-stereo: Adaptive frequency information selection for stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19701–19710, 2024. 6

[13] Zhuofeng Wu, Doehyung Lee, Zihua Liu, Kazunori Yoshizaki, Yusuke Monno, and Masatoshi Okutomi. Disparity estimation using a quad-pixel sensor. *arXiv preprint arXiv:2409.00665*, 2024. 3, 4, 5, 6

[14] Shumian Xin, Neal Wadhwa, Tianfan Xue, Jonathan T Barron, Pratul P Srinivasan, Jiawen Chen, Ioannis Gkioulekas, and Rahul Garg. Defocus map estimation and deblurring from a single dual-pixel image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2228–2238, 2021. 1

[15] Gangwei Xu, Xianqi Wang, Xiaohuan Ding, and Xin Yang. Iterative geometry encoding volume for stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21919–21928, 2023. 6

[16] Haoliang Zhao, Huizhou Zhou, Yongjun Zhang, Jie Chen, Yitong Yang, and Yong Zhao. High-frequency stereo matching network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1327–1336, 2023. 6