Supplementary material

This is a supplementary material for the paper, *Facial Depth and Normal Estimation using Single Dual-Pixel Camera*. We will further describe details: ground truth acquisition process of our facial DP dataset (Sec. A), concrete description of evaluation metrics (Sec. B), precise algorithm of our Adaptive Normal Module (Sec. C), and generalization experiments using our network and dataset (Sec. D).

A Ground-Truth Acquisition

In this section, we describe the precise ways of ground truth depth acquisition processes from our hardware setup. These processes consist of three subprocesses: (1) depth from structured light (Sec. A.1) (2) multi-view depth refinement (Sec. A.2) (3) normal-assisted depth refinement (Sec. A.4). To this end, we explain detail of our process to acquire parameters of Eq. 1 in manuscript.

A.1 Depth from Structured Light

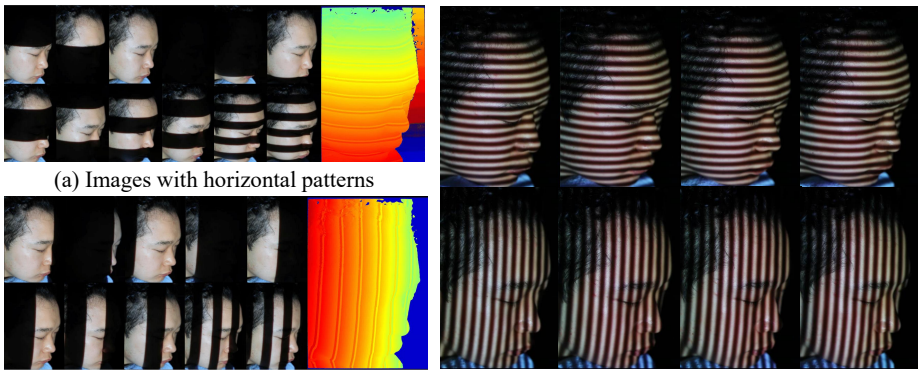
Under the carefully designed hardware setup, we start to acquire initial ground truth depth maps from structured light. We would like to briefly explain detail processes to acquire unwrapped phase images from coded patterns. As in [24,23], we use 12 horizontal patterns and 10 vertical patterns consisting of 6-bit inverse gray-code (Fig. 13-(a),(b)) to get unwrapped phase images, and 8 phase-shifting patterns [19,55] (Fig. 13-(c)) for phase correction [30]. These pattern images are used to acquire two (horizontal/vertical) unwrapped phase images per camera view. Then, we estimate dense correspondences based on a standard phase unwrapping method [55]. After that, consistency between multi-view unwrapped phase images' intensity is used to get high quality of facial depth following [23].

A.2 Multi-View Depth Refinement

Structured light can give us high-quality facial geometry that can be regarded as the ground-truth depth maps. There still exists outliers that mainly comes from small movement of face while capturing different coded patterns. To resolve this problem, we check the visibility of acquired points in perspective view of each camera and remove outliers of gathered point clouds with considering neighbor points. Finally, we check each points' multi-view photometric/depth consistency (each point should be visible from more than two sampled views) to determine whether the point is inlier or not. Results in Fig. 14 shows that this refinement process can reduce outliers effectively without loosing inliers.

A.3 Surface Normal from Photometric Stereo

In this section, we will explain our progress to get good quality of normal map to be used for supervised signal in training step. We use a photometric stereo



(a) Images with horizontal patterns

(b) Images with vertical patterns

(c) Images with phase-shifting patterns

Fig. 13. Gray-code patterns and decoding process used in Structured-Light based ground-truth acquisition. (a) Images captured with 12 horizontal patterns and acquired unwrapped phase image. (b) Images captured with 10 vertical patterns and acquired unwrapped phase image. (c) Images captured with phase-shifting patterns to refine acquired unwrapped phase images.

(a) Before refinement

(b) After multi-view refinement

Fig. 14. Results of multi-view depth refinement. We visualize point clouds (a) before refinement and (b) after multi-view refinement.

as reported in [49] to estimate surface normal of the subjects' face. First, we capture multiple images with light from varying directions. We calibrate the lighting directions using a chrome ball and estimate them as follows:

$$\begin{aligned} \mathbf{L} = 2(\mathbf{n}\cdot\mathbf{R})\mathbf{N} - \mathbf{R}, \text{where } \mathbf{R} = (0,0,1)^\intercal, \mathbf{n} = \frac{1}{r}(n_x, n_y, n_z), \\ n_x = h_x - c_x, n_y = h_y - c_y, n_z = \sqrt{r^2 - n_x^2 - n_y^2}, \end{aligned} \tag{6}$$

where c and h are the center of a chrome ball, and the center of specular reflection, respectively. r is the norm of the normal vector (n_x, n_y, n_z). By finding the distance between the centers c and h, the lighting direction $\mathbf{L} = (L_x, L_y, L_z)$ can be calculated using Eq. 6.

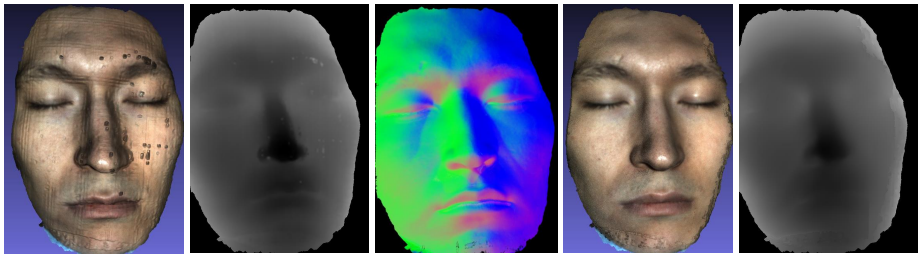


Fig. 15. Visualization of an example face of the normal-guided depth refinement. (a) 3D mesh and (b) depth map before refinement. (c) An estimated normal map. (d) Refined mesh and (e) depth map after normal-guided depth refinement.

Then, we compute the surface normal of the subject by solving a large over-constrained linear system as below:

$$\begin{bmatrix} \mathbf{N}_1 \\ \dots \\ \mathbf{N}_p \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1 \\ \dots \\ \mathbf{I}_p \end{bmatrix} [L_x, L_y, L_z]^\intercal, \quad p \in \mathbf{P}, \tag{7}$$

Finally, we follow outlier-robust scheme [48,27] to reject non-Lambertian observations by regarding them as outliers.

A.4 Normal-guided Depth Refinement

We further continue to remove outliers and improve the quality of depth by constraining the surface gradient with given normal map from Sec. A.3, which is orthogonal to the photometric normal while keeping its original position as possible. For this, we formulate an energy function consisting of two error terms: the position error E_d and the normal error E_n:

$$\begin{aligned} E &= \lambda E_d + (1-\lambda) E_n, \\ E_d &= \left\| \mathbf{X}_p - \mathbf{X}_p^m \right\|, \\ E_n &= \sum_p \left(\left[T_x \left(\mathbf{X}_p \right) \cdot \mathbf{N}_p \right]^2 + \left[T_y \left(\mathbf{X}_p \right) \cdot \mathbf{N}_p \right]^2 \right) \end{aligned} \tag{8}$$

where $\mathbf{X}_p$, $T_{x,y}(\mathbf{X}_p)$ are a 3D point and its surface gradient in the x, y-direction, respectively. λ is the balancing term between the position and the normal errors. The entire minimization is also formulated as an over-constrained linear system to be solved by least squares as described in [49].

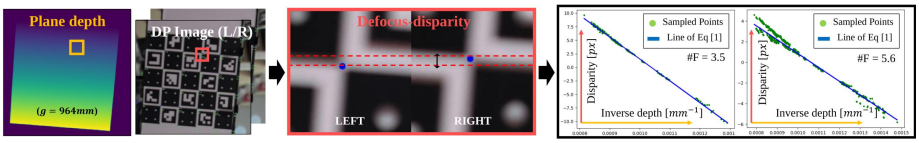
As shown in the Fig. 15, the refinement process effectively removes outliers and improves the overall quality of facial geometry by alleviating the line-artifact and noisy 3D points.

A.5 Conversion from Disparity to Metric Depth

Given the estimated defocus-disparity from our proposed network, StereoDP-Net, we provide exact conversion between the disparity and the metric depth

(a) DP input (b) Disparity / Surface Normal (c) Correct conversion (d) Incorrect conversion

Fig. 16. Facial reconstruction from given facial DP image. Given DP input of (a), our proposed StereoDPNet estimates disparity map and surface normal in (b). We show the reconstructed point cloud using correct/incorrect conversion of Sec. A.5 in (c), (d) to show the importance of our calibration process.



(a) Acquisition of depth and defocus-disparity from calibration process. (b) Find parameters of Eq [1] by line fitting.

Fig. 17. Defocus-disparity to metric depth. (a) We first acquire defocus-disparity and depth obtained by a plane homography. (b) Using the acquired depth and disparity, we find parameters of Eq. 1.

in Sec. 4.3 in the manuscript. As shown in Fig. 16, finding the correct relationship is critical to facial 3D reconstruction since the wrong conversion can result twisted shape. Here, we explain our calibration process in details that covers conversion between a defocus-disparity and a metric-scale depth.

In Fig. 17 (a), we compute the corresponding pair of points of left/right DP images. In particular, we adopt the saddle point refinement method [25] that is robust to defocus blur in DP images. By doing so, we obtain plane depth and defocus-disparity at each point.

In Fig. 17 (b), following Eq. 1, we use the linear relation between the obtained inverse depth and defocus-disparity measurements. To find calibration parameters, we first compute the bias $B(L, f, g)$ and the slope $A(L, f, g)$ through least-square optimization. Then, we obtain the focus distance $g{=}-\frac{A(L,f,g)}{B(L,f,g)}$. Using focal length f and F number that are pre-defined by the lens condition, we calculate the aperture L. α is acquired from the slope $A(L, f, g)$. Finally, we get all the calibration parameters of Eq. 1.

B Evaluation Metrics

Depth Metrics. Previous studies [16,51] predict depth with affine ambiguity. Therefore, their methods only provide experimental results with affine invariant metrics. Following [16], we measure the quality of affine transformed depth as:

- Affine invariant metrics
  - AIWE(p) : $\min_{a,b} \left(\frac{\sum_{u=1}^{W} \sum_{v=1}^{H} \left| d_{u,v} - (a\hat{d}_{u,v} + b) \right|^p}{|H \cdot W|} \right)^{1/p}$
  - WMAE = AIWE(1)
  - WRMSE = AIWE(2)

Furthermore, based on our calibration parameters (Sec. A.5), we are able to measure the accuracy of absolute-scale depth[4] as:

- Absolute metrics
  - RMSE : $\sqrt{\frac{1}{|H \cdot W|} \sum_{u=1}^{W} \sum_{v=1}^{H} \left| Z_{u,v} - \hat{Z}_{u,v} \right|^2}$
  - AbsRel : $\frac{1}{|H \cdot W|} \sum_{u=1}^{W} \sum_{v=1}^{H} \left| \left(Z_{u,v} - \hat{Z}_{u,v} \right) / Z_{u,v} \right|$
  - MAE : $\frac{1}{|H \cdot W|} \sum_{u=1}^{W} \sum_{v=1}^{H} \left| Z_{u,v} - \hat{Z}_{u,v} \right|$
  - δ^i : $\frac{1}{|H \cdot W|} \sum_{u=1}^{W} \sum_{v=1}^{H} \left(\max \left(\frac{Z_{u,v}}{\hat{Z}_{u,v}}, \frac{\hat{Z}_{u,v}}{Z_{u,v}} \right) < \tau^i \right)$

where $\hat{Z}$ denotes estimated depth and Z denotes ground-truth depth. Here, we used τ as 1.01 where $i \in \{1, 2, 3\}$.

Normal Metrics. Following [38], we use Mean Angular Error (MAE) and Root Mean Square Angular Error (RMSAE) as:

- Normal metrics
  - MAE : $\frac{1}{|H \cdot W|} \sum_{u=1}^{W} \sum_{v=1}^{H} \arccos \left(\mathbf{n}_{u,v} \cdot \hat{\mathbf{n}}_{u,v} \right)$
  - RMSAE : $\sqrt{\frac{1}{|H \cdot W|} \sum_{u=1}^{W} \sum_{v=1}^{H} \arccos \left(\mathbf{n}_{u,v} \cdot \hat{\mathbf{n}}_{u,v} \right)^2}$

[4] http://www.cvlibs.net/datasets/kitti/

C Details of Adaptive Normal Module

In this paper, we propose two sub-modules for predicting depth and surface normal from a single pair of facial DP images. Here, we describe details of surface sampling layer of our proposed ANM in Alg. 1, and Alg. 2.

Algorithm 1 Surface Sampling in ANM

Require: Aggregated Cost Volume, $C_{\mathrm{A}} \in \mathbb{R}^{C \times M \times H \times W}$
Inferred disparity, $\hat{d} \in \mathbb{R}^{H \times W}$
Number of sampled neighbors, P=4
Number of hypothesis planes in C_{A}, M=8

procedure SurfaceSample(C_{A}, $\hat{d}$, P, M)
  $\tilde{d}_{\mathrm{P}} \leftarrow$Convert-Disparity-to-VolumeIndex($\hat{d}, M$) ▷ Alg. 2
  $C_S \leftarrow$Sample-P-Closest-Neighbors($C_A, \tilde{d}_{\mathrm{P}}, P, M$)
  return Sampled volume $C_S \in \mathbb{R}^{C \times P \times H \times W}$
end procedure

Algorithm 2 Ray Sampling in Volume

Require: Inferred disparity $\hat{d} \in \mathbb{R}^{H \times W}$
Number of hypothesis planes in C_{A}, M=8
$d_{\min} = d^0$, $d_{\max} = d^m$ (Eq. 2 of the manuscript)
VolumeIndex $\tilde{d}_{\mathrm{P}} \in \mathbb{R}^{3 \times H \times W}$

procedure Convert-Disparity-to-VolumeIndex($\hat{d}$, M)
  for (u, v) in $\hat{d}$ **do**
    $\tilde{d}_{\mathrm{P}(u,v)} \leftarrow \left(u, v, \frac{\hat{d}_{(u,v)} - d_{\min}}{d_{\max} - d_{\min}} \cdot M\right)$
  end for
  return VolumeIndex $\tilde{d}_{\mathrm{P}} \in \mathbb{R}^{3 \times H \times W}$
end procedure

D Supplementary Results

D.1 Comparison of DPNet with the Original

Since there is no public code for DPNet [16], we re-implement DPNet [16] by ourselves following their description. To verify our implementation, we train our DPNet and measure the performance on their dataset [16]. Although there is a little performance drop with our implementation compared to the reported performance in the original paper [16], we show that our implemented model has similar performance with the original implementation in Table 5.

Method	Affine error metric ↓		
	AIWE(1)	AIWE(2)	1 - ρ
DPNet (reported in [16])	0.0581	0.0735	0.827
DPNet (reimplemented)	0.073	0.09	0.883

Table 5. Comparison of DPNet reported in [16] and re-implemented by ours. We measure the performance of our re-implemented DPNet on their dataset [16] and compare with the reported performance in [16]. Note that we don't use their affine invariant loss to purely verify the performance of the model.

D.2 Application : Face Relighting

Although our main goal is to estimate the depth and normal from DP images, we introduce naive methods for one of the applications, face relighting, that can be the baseline for future works.

Face Relighting. Given the reference images and the surface normals from StereoDPNet, we generate relighted images using a Ratio Image-based method [75]. The target spherical harmonic lightings are randomly sampled and the lighting directions of reference images are inferred with SfSNet [58]. The results are shown in Fig. 18.

D.3 Real-World Results

We show additional real-world results with unmet environment to demonstrate our method and dataset's generality in Fig. 20 and in Fig. 21 similar to Fig. 8 of manuscript. We also note that our method is able to apply with various camera parameters (focus distance from 1.0m to 1.5m and F-number from 2.0 to 7.1). We also demonstrate that our method can deal with various facial expressions in Fig. 19.

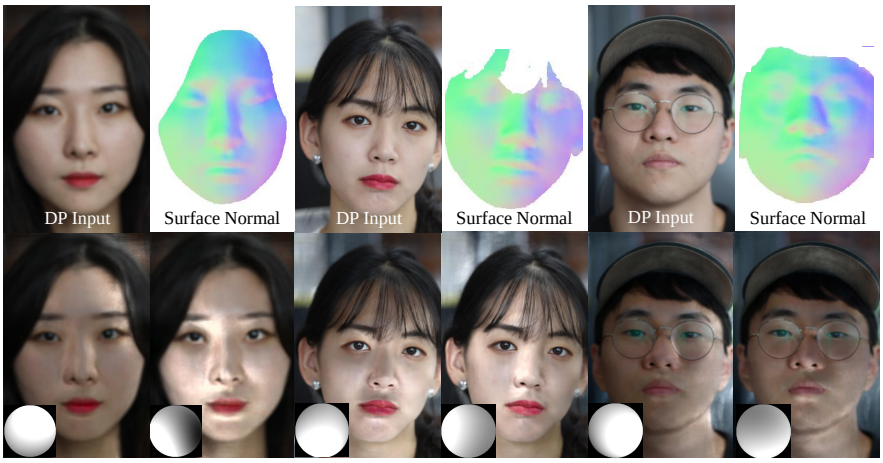


Fig. 18. Face relighting from our estimated surface normal. We display the original DP image, estimated normal map from StereoDPNet, and two different relit images with sampled lighting directions.

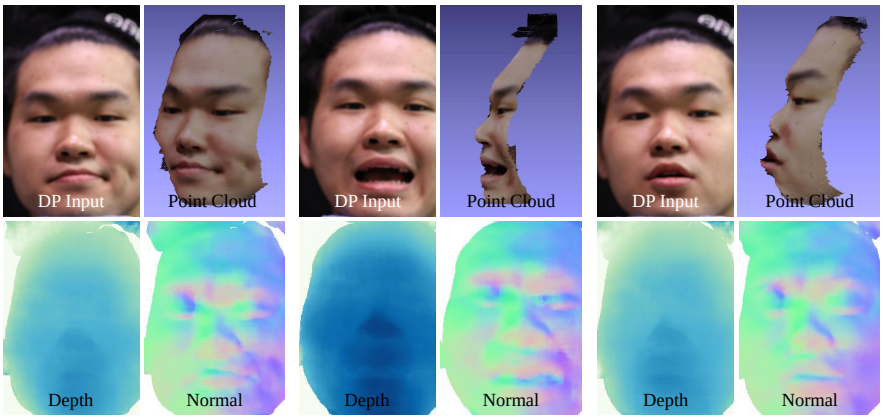


Fig. 19. Depth and Normal with facial expressions. We demonstrate that our method can cover face with various facial expressions thanks to our carefully designed facial dataset.

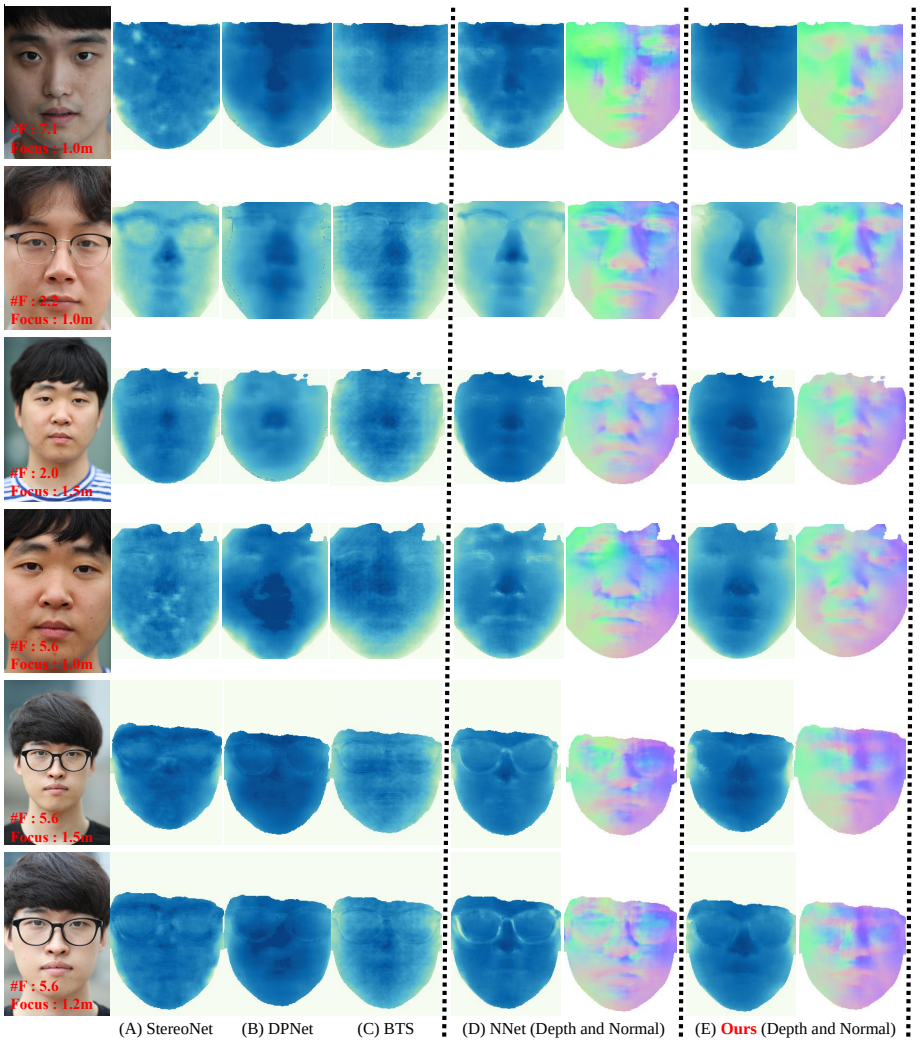


Fig. 20. Real-world results. More real-world results with captured camera settings similar to Fig. 8.

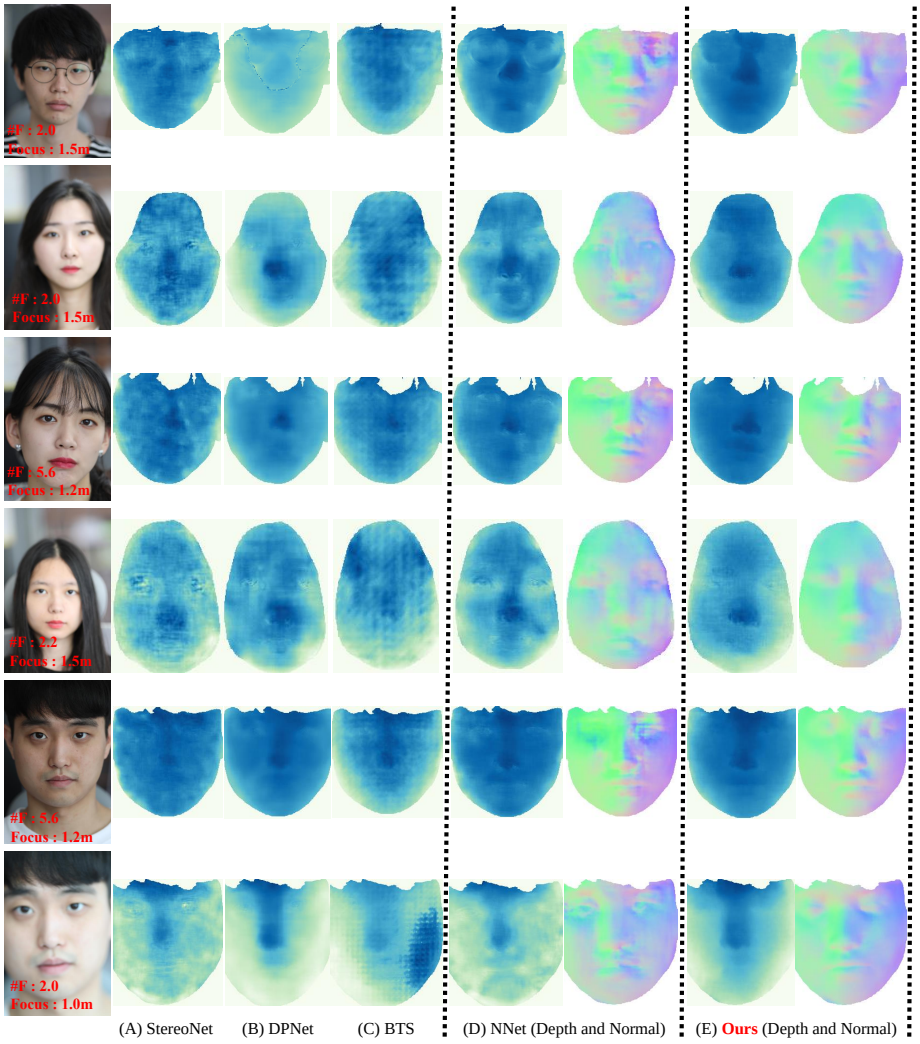


Fig. 21. Real-world results. More real-world results with captured camera settings similar to Fig. 8.

References

1. Aanæs, H., Jensen, R.R., Vogiatzis, G., Tola, E., Dahl, A.B.: Large-scale data for multiple-view stereopsis. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) pp. 1–16 (2016)
2. Abuolaim, A., Brown, M.S.: Defocus deblurring using dual-pixel data. In: Proceedings of the European conference on computer vision (ECCV). pp. 111–126. Springer (2020)
3. Abuolaim, A., Delbracio, M., Kelly, D., Brown, M.S., Milanfar, P.: Learning to reduce defocus blur by realistically modeling dual-pixel data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2289–2298 (2021)
4. Apple: Apple iphone 11 pro. `https://www.apple.com/iphone-11-pro/` (2019), accessed: 2019-09-20
5. ARCore: Augmented faces. `https://developers.google.com/ar/develop/java/augmented-faces` (2019), accessed: 2019-12-18
6. Bai, Z., Cui, Z., Rahim, J.A., Liu, X., Tan, P.: Deep facial non-rigid multi-view stereo. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5850–5860 (2020)
7. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques. pp. 187–194 (1999)
8. Boss, M., Jampani, V., Kim, K., Lensch, H., Kautz, J.: Two-shot spatially-varying brdf and shape estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3982–3991 (2020)
9. Chang, J.R., Chen, Y.S.: Pyramid stereo matching network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
10. Chen, C.H., Zhou, H., Ahonen, T.: Blur-aware disparity estimation from defocus stereo images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 855–863 (2015)
11. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV) (2018)
12. Chen, W., Mirdehghan, P., Fidler, S., Kutulakos, K.N.: Auto-tuning structured light by optical stochastic gradient descent. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
13. Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., Tong, X.: Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2019)
14. Feng, Y., Wu, F., Shao, X., Wang, Y., Zhou, X.: Joint 3d face reconstruction and dense alignment with position map regression network. In: Proceedings of the European conference on computer vision (ECCV) (2018)
15. Galaxy: Samsung galaxy s10. `https://www.samsung.com/us/mobile/galaxy-s10/` (2019), accessed: 2019-03-08
16. Garg, R., Wadhwa, N., Ansari, S., Barron, J.T.: Learning single camera depth estimation using dual-pixels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019)

17. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
18. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012)
19. Geng, J.: Structured-light 3d surface imaging: a tutorial. Advances in Optics and Photonics **3**(2), 128–160 (2011)
20. Google: Google photos: One year, 200 million users, and a whole lot of selfies. `https://blog.google/products/photos/google-photos-one-year-200-million/` (2016), accessed: 2016-05-27
21. Google: More controls and transparency for your selfies. `https://blog.google/outreach-initiatives/digital-wellbeing/more-controls-selfie-filters/` (2020), accessed: 2020-10-01
22. Guo, J., Zhu, X., Yang, Y., Yang, F., Lei, Z., Li, S.Z.: Towards fast, accurate and stable 3d dense face alignment. In: Proceedings of the European conference on computer vision (ECCV). pp. 152–168. Springer (2020)
23. Ha, H., Oh, T.H., Kweon, I.S.: A multi-view structured-light system for highly accurate 3d modeling. In: International Conference on 3D Vision (3DV) (2015)
24. Ha, H., Park, J., Kweon, I.S.: Dense depth and albedo from a single-shot structured light. In: International Conference on 3D Vision (3DV). pp. 127–134 (2015)
25. Ha, H., Perdoch, M., Alismail, H., So Kweon, I., Sheikh, Y.: Deltille grids for geometric camera calibration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5344–5352 (2017)
26. Han, Y., Lee, J.Y., So Kweon, I.: High quality shape from a single rgb-d image under uncalibrated natural illumination. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2013)
27. Hernandez, C., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(3), 548–554 (2008)
28. Hu, P., Ramanan, D.: Finding tiny faces. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 951–959 (2017)
29. Im, S., Ha, H., Choe, G., Jeon, H.G., Joo, K., Kweon, I.S.: High quality structure from small motion for rolling shutter cameras. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2015)
30. Je, C., Lee, S.W., Park, R.H.: Color-phase analysis for sinusoidal structured light in rapid range imaging (2004)
31. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H.: Large scale multi-view stereopsis evaluation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
32. Jeon, H.G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.W., Kweon, I.S.: Depth from a light field image with learning-based matching costs. IEEE transactions on pattern analysis and machine intelligence **41**(2), 297–310 (2018)
33. Jeon, H.G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.W., So Kweon, I.: Accurate depth map estimation from a lenslet light field camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
34. Keselman, L., Iselin Woodfill, J., Grunnet-Jepsen, A., Bhowmik, A.: Intel realsense stereoscopic depth cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2017)
35. Khamis, S., Fanello, S., Rhemann, C., Kowdle, A., Valentin, J., Izadi, S.: Stereonet: Guided hierarchical refinement for real-time edge-aware depth prediction. In: Pro-

ceedings of the European Conference on Computer Vision (ECCV). pp. 573–590 (2018)
36. Kinect2: Kinect for windows sdk 2.0. `https://developer.microsoft.com/en-us/windows/kinect/` (2014), accessed: 2014-10-21
37. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
38. Kusupati, U., Cheng, S., Chen, R., Su, H.: Normal assisted stereo depth estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
39. Lattas, A., Moschoglou, S., Gecer, B., Ploumpis, S., Triantafyllou, V., Ghosh, A., Zafeiriou, S.: Avatarme: Realistically renderable 3d facial reconstruction" in-the-wild". In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 760–769 (2020)
40. Lee, J.H., Han, M.K., Ko, D.W., Suh, I.H.: From big to small: Multi-scale local planar guidance for monocular depth estimation. arXiv preprint arXiv:1907.10326 (2019)
41. Liang, J., Tu, H., Liu, F., Zhao, Q., Jain, A.K.: 3d face reconstruction from mugshots: Application to arbitrary view face recognition. Neurocomputing **410**, 12–27 (2020)
42. Lichy, D., Wu, J., Sengupta, S., Jacobs, D.W.: Shape and material capture at home. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6123–6133 (2021)
43. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
44. Liu, F., Zhao, Q., Liu, X., Zeng, D.: Joint face alignment and 3d face reconstruction with application to face recognition. IEEE transactions on pattern analysis and machine intelligence **42**(3), 664–678 (2018)
45. Long, X., Lin, C., Liu, L., Li, W., Theobalt, C., Yang, R., Wang, W.: Adaptive surface normal constraint for depth estimation. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
46. Long, X., Liu, L., Theobalt, C., Wang, W.: Occlusion-aware depth estimation with adaptive normal constraints. In: Proceedings of the European conference on computer vision (ECCV). pp. 640–657. Springer (2020)
47. Luo, H., Nagano, K., Kung, H.W., Xu, Q., Wang, Z., Wei, L., Hu, L., Li, H.: Normalized avatar synthesis using stylegan and perceptual refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11662–11672 (2021)
48. Mukaigawa, Y., Ishii, Y., Shakunaga, T.: Analysis of photometric factors based on photometric linearization. JOSA A **24**(10), 3326–3334 (2007)
49. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3d geometry. ACM Transactions on Graphics (ToG) **24**(3), 536–543 (2005)
50. Pan, L., Chowdhury, S., Hartley, R., Liu, M., Zhang, H., Li, H.: Dual pixel exploration: Simultaneous depth estimation and image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4340–4349 (June 2021)
51. Punnappurath, A., Abuolaim, A., Afifi, M., Brown, M.S.: Modeling defocus-disparity in dual-pixel sensors. In: 2020 IEEE International Conference on Computational Photography (ICCP) (2020)

52. Qi, X., Liao, R., Liu, Z., Urtasun, R., Jia, J.: Geonet: Geometric neural network for joint depth and surface normal estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 283–291 (2018)
53. Qiu, J., Cui, Z., Zhang, Y., Zhang, X., Liu, S., Zeng, B., Pollefeys, M.: Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
54. Richardson, E., Sela, M., Or-El, R., Kimmel, R.: Learning detailed face reconstruction from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1259–1268 (2017)
55. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern Recognition **43**(8), 2666–2680 (2010)
56. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). vol. 1 (2003)
57. Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: Proceedings of the European conference on computer vision (ECCV) (2016)
58. Sengupta, S., Kanazawa, A., Castillo, C.D., Jacobs, D.W.: Sfsnet: Learning shape, reflectance and illuminance of facesin the wild'. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
59. Shang, J., Shen, T., Li, S., Zhou, L., Zhen, M., Fang, T., Quan, L.: Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry consistency. In: Proceedings of the European conference on computer vision (ECCV). pp. 53–70. Springer (2020)
60. Shi, B., Wu, Z., Mo, Z., Duan, D., Yeung, S.K., Tan, P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
61. Silberman, N., Fergus, R.: Indoor scene segmentation using a structured light sensor. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) - Workshop on 3D Representation and Recognition (2011)
62. Song, G., Zheng, J., Cai, J., Cham, T.J.: Recovering facial reflectance and geometry from multi-view images. Image and Vision Computing **96**, 103897 (2020)
63. Tran, L., Liu, X.: Nonlinear 3d face morphable model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7346–7355 (2018)
64. Wadhwa, N., Garg, R., Jacobs, D.E., Feldman, B.E., Kanazawa, N., Carroll, R., Movshovitz-Attias, Y., Barron, J.T., Pritch, Y., Levoy, M.: Synthetic depth-of-field with a single-camera mobile phone. ACM Transactions on Graphics (ToG) **37**(4), 1–13 (2018)
65. Wu, F., Bao, L., Chen, Y., Ling, Y., Song, Y., Li, S., Ngan, K.N., Liu, W.: Mvf-net: Multi-view 3d face morphable model regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 959–968 (2019)
66. Wu, S., Rupprecht, C., Vedaldi, A.: Unsupervised learning of probably symmetric deformable 3d objects from images in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)

67. Wu, X., Zhou, J., Liu, J., Ni, F., Fan, H.: Single-shot face anti-spoofing for dual pixel camera. IEEE Transactions on Information Forensics and Security **16**, 1440–1451 (2020)
68. Xin, S., Wadhwa, N., Xue, T., Barron, J.T., Srinivasan, P.P., Chen, J., Gkioulekas, I., Garg, R.: Defocus map estimation and deblurring from a single dual-pixel image. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
69. Xu, H., Zhang, J.: Aanet: Adaptive aggregation network for efficient stereo matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1959–1968 (2020)
70. Yang, F., Wang, J., Shechtman, E., Bourdev, L., Metaxas, D.: Expression flow for 3d-aware face component transfer. In: ACM SIGGRAPH 2011 papers, pp. 1–10 (2011)
71. Ying, X., Wang, L., Wang, Y., Sheng, W., An, W., Guo, Y.: Deformable 3d convolution for video super-resolution. IEEE Signal Processing Letters **27**, 1500–1504 (2020)
72. Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., Zhao, G.: Deep learning for face anti-spoofing: A survey. arXiv preprint arXiv:2106.14948 (2021)
73. Zhang, F., Prisacariu, V., Yang, R., Torr, P.H.: Ga-net: Guided aggregation net for end-to-end stereo matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 185–194 (2019)
74. Zhang, Y., Wadhwa, N., Orts-Escolano, S., Häne, C., Fanello, S., Garg, R.: Du 2 net: Learning depth estimation from dual-cameras and dual-pixels. In: Proceedings of the European conference on computer vision (ECCV). pp. 582–598. Springer (2020)
75. Zhou, H., Hadap, S., Sunkavalli, K., Jacobs, D.W.: Deep single-image portrait relighting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019)
76. Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., Stamminger, M., Nießner, M., Theobalt, C.: State of the art on monocular 3d face reconstruction, tracking, and applications. In: Computer Graphics Forum. vol. 37, pp. 523–550. Wiley Online Library (2018)