This WACV paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the accepted version; the final published version of the proceedings is available on IEEE Xplore.

Polarimetric PatchMatch Multi-View Stereo

Jinyu Zhao Jumpei Oishi Yusuke Monno Masatoshi Okutomi Tokyo Institute of Technology

{jzhao, joishi, ymonno}@ok.sc.e.titech.ac.jp, mxo@ctrl.titech.ac.jp

Abstract

PatchMatch Multi-View Stereo (PatchMatch MVS) is one of the popular MVS approaches, owing to its balanced accuracy and efficiency. In this paper, we propose Polarimetric PatchMatch multi-view Stereo (PolarPMS), which is the first method exploiting polarization cues to PatchMatch MVS. The key of PatchMatch MVS is to generate depth and normal hypotheses, which form local 3D planes and slanted stereo matching windows, and efficiently search for the best hypothesis based on the consistency among multi-view images. In addition to standard photometric consistency, our PolarPMS evaluates polarimetric consistency to assess the validness of a depth and normal hypothesis, motivated by the physical property that the polarimetric information is related to the object's surface normal. Experimental results demonstrate that our PolarPMS can improve the accuracy and the completeness of reconstructed 3D models, especially for texture-less surfaces, compared with state-ofthe-art PatchMatch MVS methods.

1. Introduction

PatchMatch Multi-View Stereo (PatchMatch MVS) [6] is one of the popular MVS approaches for its good balance in reconstruction accuracy and efficiency [12, 18, 22, 28]. PatchMatch MVS introduces the idea of PatchMatch [5] to realize an accurate and fast global search for multi-view correspondences. Specifically, PatchMatch MVS methods generally estimate the depth and normal for each pixel of each reference view as follows: (i) Initialize the depth and normal randomly; (ii) Generate depth and normal hypotheses by random generation, perturbation of current estimations, and the propagation considering an adjacent pixel; (iii) Evaluate multi-view consistency to select the best depth and normal hypothesis; (iv) Perform the processes (ii) and (iii) for all the pixels in the scanning order of rowmajor/column-major and iterate the whole process until sufficient depth and normal results are derived.

In the above processes, a generated depth and normal hypothesis for a pixel of a reference view is used to form a



Figure 1. For the evaluation of a depth and normal hypothesis, our method evaluates standard photometric consistency and additional polarimetric consistency, which poses a cost to the inconsistency between the observed angle-of-polarization (AoP) and the azimuth angle of the normal hypothesis.

local 3D plane, as shown in the top part of Fig. 1. This local plane is projected to a source view, which forms a slanted patch for stereo matching. To evaluate the reliability of the depth and normal hypothesis, PatchMatch MVS methods evaluate the photometric consistency (color similarity) between the corresponding patches in the reference and the source views. However, since both the depth and the normal are evaluated based on color textures, it remains challenging to accurately estimate them for texture-less regions.

In this paper, we propose Polarimetric PatchMatch multi-view Stereo (PolarPMS), which is the first method exploiting polarization cues to PatchMatch MVS. Our PolarPMS is motivated by the physical property that the angle-of-polarization (AoP) of reflected light from an object is related to the azimuth angle calculated from the object's surface normal [8, 27]. Based on this, our PolarPMS evaluates polarimetric consistency, which is the consistency between the observed AoP and the azimuth angle from the normal, as shown in the bottom part of Fig. 1. Compared with exist-

ing PatchMatch MVS methods solely based on photometric consistency, our PolarPMS utilizing additional polarimetric consistency mainly has two benefits: (i) Since the AoP is a direct cue to assess the surface normal, the accuracy of the estimated normal map is significantly improved. Furthermore, the improved normal estimation can derive improved depth estimation by effectively incorporating depthnormal consistency evaluation. (ii) Since the AoP is not influenced by the surface texture, the incorporation of polarimetric consistency significantly improves the completeness for texture-less regions. To summarize, the main contributions of this work are as follows:

- We propose PolarPMS, which is the first PatchMatch MVS method exploiting polarization information.
- We introduce polarimetric consistency based on AoP, which acts as a direct cue to assess the surface normal regardless of the surface texture, to select the best depth and normal hypothesis in PatchMatch MVS.
- We introduce depth-normal consistency to improve the depths by utilizing improved normals with AoP.
- We experimentally validate the effectiveness of our PolarPMS, which can derive improved 3D models, especially for texture-less surfaces.
- We make the source code publicly available at http://www.ok.sc.e.titech.ac.jp/res/
 PolarPMS/.

2. Related Works

2.1. PatchMatch Multi-View Stereo

PatchMatch MVS was originally proposed in [6] and its improved versions have been actively proposed [12, 18, 22, 28]. Inspired by the PatchMatch search algorithm originally designed for image patches [5], PatchMatch MVS realizes an accurate and fast global search of the best 3D plane hypothesis that is the most consistent with multi-view input images. However, as common to most other MVS approaches, it remains challenging to derive accurate depth and normal maps for texture-less regions, because Patch-Match MVS methods rely on photometric consistency evaluation to select the best hypothesis.

Some recent PatchMatch MVS methods attempt to address this challenge. TAPA-MVS [16] represents textureless regions as superpixels and fits a plane for each superpixel. ACMM [24] applies a multi-scale guided approach, where texture-less regions are considered as better textured when the original images are downsampled. ACMP [25] and ACMMP [23] combine planar priors to ACMM, which can provide additional geometric constraints to texture-less regions. Even though the plane-based approaches work well for planer scenes, such as buildings and indoor rooms, they may lose surface details if general objects are targeted.

2.2. Multi-View Polarimetric Reconstruction

It is known that the azimuth angle and the zenith angle of the object's surface normal are related to AoP and degree-of-polarization (DoP) of reflected light from the object [8, 27]. Based on these physical properties, many polarimetric 3D reconstruction methods have been proposed in a single-view setting [19], a two-view stereo setting [4, 11], or a multi-view setting [8, 26, 27].

Polarimetric MVS [8] and Polarimetric MVIR [27] are two closely related methods to our PolarPMS, which adopt multi-view polarization images as the inputs. Polarimetric MVS exploits AoP information to decide the directions to propagate the sparse depths obtained by SfM. While Polarimetric MVS derives dense depth maps even for texture-less regions, the depth propagation is performed in a view-byview manner without checking multi-view consistency regarding the polarization, leading to limited accuracy.

Polarimetric MVIR and our PolarPMS utilize a similar observation regarding the relationship between the AoP and the azimuth angle in multi-view consistency. However, they play different roles in the 3D reconstruction steps. PolarPMS is an MVS method that builds a dense point cloud from scratch and uses AoP information to search for the best 3D plane hypothesis in the PatchMatch MVS framework. In contrast, Polarimetric MVIR is a refinement method based on a reasonable initial shape from MVS and uses AoP information for global mesh optimization. Thus, these methods can be applied to each reconstruction step in combination.

2.3. Deep-Learning-Based Methods

Deep-learning-based methods also have been emerging for PatchMatch MVS [10, 13, 21] and multi-view polarimetric 3D reconstruction [7, 20]. While these methods demonstrate high potential and performance on their benchmarks, a large amount of training data is still hard to obtain especially for polarization images, for which simulation tools and open datasets are very limited. While a recent method of [9] based on neural radiance fields does not require the training data, it tends to generate over-smoothed surfaces due to the convolutional nature of the network. Since our PolarPMS builds a 3D model as in a classical MVS approach, it falls into a much different category from those learning-based methods.

3. Polarimetric PatchMatch Multi-View Stereo

3.1. Depth and Normal Estimation Overview

Our PolarPMS uses multiple polarization images taken from different viewpoints, and camera poses are derived from SfM [17]. Figure 2 shows the flow of depth and normal estimation for a reference view, which is one of the input views. Firstly, the depth and normal for each pixel are randomly initialized. Then, based on the initial/current



Figure 2. Flow of depth and normal estimation for a reference view in our PolarPMS. According to the PatchMatch algorithm, the depth and normal hypothesis generation and evaluation are performed pixel-wisely in the scanning order of row-major/reverse-row-major/column-major/reverse-column-major at each iteration. We utilize polarimetric information to evaluate the depth and normal hypotheses.

depth and normal, seven depth and normal hypotheses are generated, as detailed later. Those hypotheses are then evaluated based on the pair-wise consistency between the reference and the source views to select the best hypothesis as the current estimation. This hypothesis generation and selection are performed pixel-wisely in the scanning order of row-major, reverse-row-major, column-major, or reversecolumn-major at each iteration until the iteration limit to obtain the final depth and normal estimations. These estimation processes are performed on every input view to obtain the depth and the normal maps of all the views.

3.2. Depth and Normal Hypothesis Generation

Given the initial/current depth and normal estimations, seven depth and normal hypotheses are generated in the same way as [18] as follows.

$$\mathcal{A} = \{ (d_l, \boldsymbol{n}_l), (d_{l-1}^{prop}, \boldsymbol{n}_{l-1}), (d_l^{rnd}, \boldsymbol{n}_l), (d_l, \boldsymbol{n}_l^{rnd}), \\ (d_l^{rnd}, \boldsymbol{n}_l^{rnd}), (d_l^{prt}, \boldsymbol{n}_l), (d_l, \boldsymbol{n}_l^{prt}) \},$$
(1)

where d_l and n_l are the current depth and normal of the pixel l, respectively. d_{l-1}^{prop} is the depth obtained by the inter-pixel propagation from the depth and normal of the adjacent pixel l-1, which is the depth hypothesis assuming a smooth surface. d_l^{rnd} and n_l^{rnd} are randomly generated depth and normal hypotheses, and d_l^{prt} and n_l^{prt} are perturbed hypotheses of the current depth and normal.

3.3. Depth and Normal Hypothesis Evaluation

3.3.1 Overall Cost Function

Here, we detail the cost function for depth and normal evaluation. The overall cost function is as follows.

$$(d_l^{opt}, \boldsymbol{n}_l^{opt}) = \arg\min_{d_l^*, \boldsymbol{n}_l^*} \Big[F_{pho}(d_l^*, \boldsymbol{n}_l^*) + \tau_{geo} \cdot F_{geo}(d_l^*, \boldsymbol{n}_l^*) + \tau_{pol} \cdot F_{pol}(d_l^*, \boldsymbol{n}_l^*) + \tau_{dep} \cdot F_{dep}(d_l^*, \boldsymbol{n}_l^*) \Big],$$

$$(2)$$

where (d_l^{opt}, n_l^{opt}) is the best hypothesis selected from \mathcal{A} for the pixel l with a minimum cost, and (d_l^*, n_l^*) represents one of the seven depth and normal hypotheses in \mathcal{A} . F_{pho} , F_{geo} , F_{pol} , and F_{dep} evaluate photometric consistency, geometric consistency, polarimetric consistency, and depth-normal consistency, respectively. τ_{geo} , τ_{pol} , and τ_{dep} are balancing weights. F_{pho} and F_{geo} are based on COLMAP [18], and we newly introduce F_{pol} and F_{dep} . The optimization problem can be solved by Ceres solver [3].

Figure 3 illustrates the geometric projections used for cost evaluation. For one depth and normal hypothesis to a considered pixel/patch in the reference image, one local 3D plane hypothesis is generated. This local 3D plane is then projected to the source image plane, by which a slanted patch for photometric consistency evaluation is derived. Similarly, the normal vector is projected to the reference and the source image planes, respectively, which are used to derive the azimuth angle of the normal for polarimetric consistency evaluation.

3.3.2 Photometric Consistency

The photometric consistency between the patch in the reference image and the corresponding patch in the source image



Figure 3. Illustrations of geometric projections used for our cost calculation.



Figure 4. Illustrations of photometric cost calculation and polarimetric cost calculation for hypothesis (d^2, n^2) of Fig. 3.

(see Fig. 4(a)) is evaluated using standard RGB images as

$$F_{pho}(d_l^*, \boldsymbol{n}_l^*) = \frac{1}{|\mathcal{S}|} \sum_{m \in \mathcal{S}} \left[1 - \sigma_l^m(d_l^*, \boldsymbol{n}_l^*) \right], \quad (3)$$

where S is the set of source views, which is adaptively selected for each pixel l of the reference image [18], and mis the index of the source view. σ_l^m is the color similarity, which is calculated as

$$\sigma_l^m(d_l^*, \boldsymbol{n}_l^*) = \frac{cov_\omega(\boldsymbol{\omega}_l, \boldsymbol{\omega}_l^m)}{\sqrt{cov_\omega(\boldsymbol{\omega}_l, \boldsymbol{\omega}_l)cov_\omega(\boldsymbol{\omega}_l^m, \boldsymbol{\omega}_l^m)}}, \quad (4)$$

where cov_{ω} is the weighted covariance of the patches, ω_l is the reference patch, and ω_l^m is the corresponding patch in the *m*-th source image.

This cost function gives a small cost for texture-less regions with similar brightness, regardless of inconsistency in geometry. To avoid such mis-evaluation, COLMAP excludes those regions, where pixel values hardly vary in the reference patch, from the estimation by assigning a maximum cost so as to be removed by later filtering. In contrast, we do not exclude those texture-less regions to improve the completeness by using polarimetric information, as will be explained in Section 3.3.4.

3.3.3 Geometric Consistency

The geometric consistency is evaluated as

$$F_{geo}(d_l^*, \boldsymbol{n}_l^*) = \frac{1}{|\mathcal{S}|} \sum_{m \in \mathcal{S}} \xi_l^m(d_l^*, \boldsymbol{n}_l^*),$$
(5)

where ξ is robustified geometric cost function:

$$\xi_l^m(d_l^*, \boldsymbol{n}_l^*) = 1 - \sigma_l^m(d_l^*, \boldsymbol{n}_l^*) + 0.5 \min(\psi_l^m(d_l^*, \boldsymbol{n}_l^*), \psi_{max}),$$
(6)

where ψ_l^m represents the distance error between the pixel coordinates of the current pixel and the forward-backward reprojected pixel using the depth value of the reference image (forward direction) and that of the source image (backward direction), and ψ_{max} is the maximum reprojection error, which is set to be three (pixel).

3.3.4 Polarimetric Consistency

Unpolarized light becomes partially polarized when it is reflected from an object's surface. It is known that the AoP of the reflected light is related to the direction of the surface normal's projection (i.e., azimuth angle). However, the relationship is not unique and depends on whether specular reflection or diffuse reflection dominates. Generally, there exist π - and $\pi/2$ -ambiguities in the relationship between the AoP (ϕ) and the azimuth angle (α) without prerequisite knowledge about lighting conditions and surface materials [27]. This means that four possible azimuth angles (ϕ , $\phi \pm \pi/2$, $\phi + \pi$) can be inferred from one observed AoP value ϕ , as illustrated in Fig. 4(b).

Based on the above physical property, we evaluate polarimetric consistency as

$$F_{pol}(d_l^*, \boldsymbol{n}_l^*) = \frac{g(\rho_l) \cdot \delta(\eta_l(\boldsymbol{n}_l^*)) + \sum_{m \in \mathcal{S}} g(\rho_l^m) \cdot \delta(\eta_l^m(d_l^*, \boldsymbol{n}_l^*))}{g(\rho_l) + \sum_{m \in \mathcal{S}} g(\rho_l^m)}$$
(7)

where ρ_l and ρ_l^m represent the DoP values of the reference pixel and the warped pixel in the *m*-th source image, respectively. $g(\rho)$ is the weighting function based on the DoP value, which is detailed later.

The symbol δ represents the function of the minimum angle difference η between the azimuth angle α of the normal hypothesis n^* and one of the four possible azimuth angles inferred from the observed AoP value ϕ , as shown in Fig. 4(b). Mathematically, this minimum angle is calculated for the reference pixel l as

$$\eta_{l} = \min(|\alpha_{l} - \phi_{l} - 2\pi|, |\alpha_{l} - \phi_{l} - 3\pi/2|, |\alpha_{l} - \phi_{l} - \pi|, |\alpha_{l} - \phi_{l} - \pi/2|, |\alpha_{l} - \phi_{l}|, (8) |\alpha_{l} - \phi_{l} + \pi/2|, |\alpha_{l} - \phi_{l} + \pi|).$$

The minimum angle for the warped pixel in the *m*-th source image η_l^m can be calculated in the same way. As the specific function δ , we adopt a concave-shaped function of [27], which assigns a higher cost if η is larger and vice versa in the range of [0, 1].

As for the weighting, since a high DoP value means high reliability of the polarization information, we assign a larger weight to the view with a larger DoP value as

$$g(\rho) = 1 - [\min(\rho, \rho_0) - \rho_0]^2 / \rho_0^2, \tag{9}$$

which increases when the DoP value ρ becomes larger and reaches the maximum value one when ρ becomes larger than the threshold ρ_0 . By using this DoP weighting, only reliable AoP values among all the reference and source views are paid attention to assess the polarimetric consistency. Furthermore, if no polarimetric information is available, which means that DoP equals zero for all the views, the polarimetric consistency is neglected, resulting in principally the same result as COLMAP relying on photometric consistency.

3.3.5 Depth-Normal Consistency

By introducing the polarimetric consistency, it is expected that the accuracy of the estimated normal is improved. Thus, we aim to affect the improved normal estimations to the depth estimations by enforcing the consistency between the normal and the depth. For this purpose, we introduce a depth-normal consistency evaluation to enforce the consistency between the estimated normal and the normal calculated from the depths in neighboring pixels. The cost function is calculated as follows.

$$F_{dep}(d_l^*, \boldsymbol{n}_l^*) = 1 - (\boldsymbol{n}_l^*)^T \cdot \boldsymbol{n}_l^{dep}(d_l^*, d_l^h, d_l^v), \quad (10)$$

where n_l^{dep} is the normal of the plane composed of the 3D points associated with the depths of the current pixel $l(d_l^*)$, its adjucent pixels in the horizontal (d_l^h) and the vertical (d_l^v) directions.

3.4. Point Cloud Generation from Multi-Views

After generating the depth and the normal maps of all input views, a point cloud is generated by fusing them. Before the fusion, we apply filtering to remove unreliable estimations. We consider that the estimation is unreliable if a reference pixel/patch x_l does not have both enough polarization cues (DoP value ρ_l) and textures (pixel intensity



(b) Sample images

Figure 5. Synthetic data generation: (a) Camera and illumination settings; (b) the examples of the synthesized images.

variation in the patch λ_l). Thus, we first filter out the pixels that do not satisfy the following conditions.

$$\mathcal{S}_l = \{ \boldsymbol{x}_l | \rho_l \ge \rho_t, \lambda_l \ge \lambda_t \}, \tag{11}$$

where ρ_t is the threshold for the DoP value, and λ_t is the variance threshold for the patch, which are set to 0.05 and 1.0 in our paper. We then follow COLMAP [18] for the filtering of remaining pixels and the fusion processes, where photometric consistency and geometric consistency are rechecked so that 3D points are generated only from consistent depth and normal estimations in multi-views.

4. Experimental Results

4.1. Comparison Using Synthetic Data

We used three CG models (Armadillo, Bunny, and Dragon) available from Stanford 3D Scanning Repository [1]. Following [27], input RGB, AoP, and DoP images were synthesized using Mitsuba 2 renderer [15], which supports a polarimetric BRDF model to simulate realistic polarization images. The synthetic data were generated using a polarized plastic material because it is only the polarized material that can add arbitrary textures. The camera poses and illumination (environment map) are shown in Fig. 5(a) and the examples of synthesized RGB, AoP, and DoP images are shown in Fig. 5(b).

We empirically set $(\tau_{geo}, \tau_{pol}, \tau_{dep})$ in Eq. (2) to (0.4, 4.0, 0.4), and ρ_0 in Eq. (9) to 0.005, and compared our PolarPMS with COLMAP [18], which is the base of



Figure 6. Numerical evaluation of depths (top) and normals (bottom) of all pixels. The vertical axis represents the proportion of the pixels whose errors are less than the threshold of the horizontal axis.



Figure 7. Depth and normal map comparison.

our method, and ACMMP [23], which is a state-of-the-art PatchMatch MVS method incorporating planer priors.

Figure 6 numerically evaluates the depth and normal estimation results of all the pixels of all the views, where *x*axis indicates the depth or normal error threshold, and *y*axis represents the proportion of the pixels with a smaller error than the threshold. We can clearly see that our PolarPMS achieves the best accuracy among compared methods with the fastest ascent speed and the highest proportion of the pixels with less error than a certain threshold. Especially, the normal accuracy is significantly improved, owing to our polarimetric consistency evaluation that directly affects the selection of reliable normal hypothesis. This improvement can be seen visually in Fig. 7, where our PolarPMS derives significantly better normal results.

We then compare 3D point cloud results derived from depth and normal maps, using the same fusion parameters for all the compared methods. The quality of the reconstructed point is evaluated using two commonly-used

Table 1. Comparisons of the average accuracy (Acc.) and completeness (Comp.) errors

		COLMAP	ACMMP	PolarPMS (Ours)	
Armadillo	# of Points	694,448	1,110,869	918,007	
	Acc.($\times 10^{-2}$)	0.622	0.679	0.522	
	Comp.($\times 10^{-2}$)	1.035	0.835	0.568	
Bunny	# of Points	385,511	1,231,652	1,192,849	
	Acc.($\times 10^{-2}$)	0.689	0.780	0.778	
	Comp.($\times 10^{-2}$)	5.373	1.483	0.867	
Dragon	# of Points	538,665	1,040,791	1,117,226	
	Acc.($\times 10^{-2}$)	0.714	0.799	0.669	
	Comp.($\times 10^{-2}$)	3.894	2.572	1.635	
Average	Acc.(×10 ⁻²)	0.675	0.753	0.656	
	Comp.($\times 10^{-2}$)	3.434	1.630	1.023	

metrics [2, 14]: "Accuracy" which is the distance from each estimated 3D point to its nearest ground-truth 3D point and "Completeness" which is the distance from each ground-truth 3D point to its nearest estimated 3D point. Table 1 summarizes the average accuracy and completeness for each model. Figures 8 and 9 show the visual comparisons for the Armadillo model and the Bunny model, respectively. Generally, COLMAP achieves good accuracy, but poor completeness, because COLMAP only focuses on the regions with rich texture and excludes the texture-less regions from the estimation due to the difficulty in photometric consistency evaluation. This is extremely conspicuous for the Bunny model of Fig. 9. ACMMP aims to reconstruct texture-less regions and exhibits the effectiveness by better completeness than COLMAP. However, ACMMP tends to lose the surface details because it assumes local planer scenes, which is not always the case, as shown in the enlarged regions of Figs. 8 and 9. In contrast, our PolarPMS achieves good performance in terms of both accuracy and completeness, thanks to our polarimetric consistency evaluation, which is effective regardless of surface textures, to derive better surface normals and depth-normal consistency evaluation to utilize the improved normals to improve depths too.

We also performed an ablation study to confirm the effectiveness of our proposed polarimetric and depth-normal consistencies. Table 2 summarizes the average pixel-wise depth and normal errors. Compared to the first column where neither polarimetric consistency nor depth-normal consistency is introduced (principally COLMAP), the second column and the third column demonstrate that the accuracy of both depth estimation and normal estimation can be improved by considering polarimetric and depth-normal consistencies, respectively. The fourth column shows that the best results can be derived by introducing polarimetric and depth-normal consistencies simultaneously.



Figure 8. Visual comparison for the Armadillo model.



Figure 9. Visual comparison for the Bunny model.

4.2. Comparison Using Real Data

We used the dataset in [27] for a toy car (56 views), a camera (32 views), and a statue (43 views). For real data, we empirically set $(\tau_{geo}, \tau_{pol}, \tau_{dep})$ to (0.4, 10.0, 0.4) and ρ_0 to 1.0, seeing the actual strengths of noise and DoPs for real scenes. The results in Fig. 10 show that COLMAP can generally reconstruct regions with relatively rich textures, but has limited performance in texture-less regions (e.g., the front window of the toy car). Compared to COLMAP, ACMMP shows good performance in overall shape reconstruction, thanks to multi-scale geometry consistency to provide additional feature points and planar prior which can provide geometric constraints for the scene (e.g., the base

Table 2. Comparisons of the average depth and normal errors

Polarimetric consistency			✓		√
Depth-normal consistency				\checkmark	\checkmark
Armadillo	Depth	0.066	0.034	0.049	0.026
	Normal (deg)	24.429	8.897	18.344	7.845
Bunny	Depth	0.338	0.155	0.346	0.103
	Normal (deg)	46.348	10.430	40.707	8.054
Dragon	Depth	0.188	0.113	0.167	0.082
	Normal (deg)	37.309	11.864	29.956	10.263
-	Depth	0.197	0.101	0.187	0.070
Average	Normal (deg)	36.029	10.397	29.669	8.721

part of the statue). However, at the same time, the planar prior is principally difficult to apply for texture-less nonplanar regions, since details of those regions are difficult to capture by fitting them to planes (e.g., the side window of the toy car). In contrast, our PolarPMS can recover textureless regions better (e.g., the front window of the car and the face of the statue), and meantime reconstruct the details better (the side window of the car, the lens part of the camera, and the bell of the statue). This demonstrates the effectiveness of our PolarPMS by utilizing polarimetric information for recovering texture-less regions and surface details.

4.3. Comparison with Polarimetric MVS

We next compare PolarPMS with Polarimetric MVS [8], which is a representative MVS method with polarization. Since the source code of Polarimetric MVS is not available, we used the input and the result data provided by the authors of [8] for comparison. Figure 11 shows the visual comparison for real car data. We can see that, compared to COLMAP which does not utilize polarization, both Polarimetric MVS and our PolarPMS reconstruct a muchcompleted 3D point cloud. While Polarimetric MVS tends to generate a very dense point cloud, it lacks accuracy and generates scattered 3D points for a single surface. This is because Polarimetric MVS uses AoP information for viewby-view depth propagation, but does not include multi-view consistency evaluation based on polarization. On the other hand, our PolarPMS generates more accurate 3D points by utilizing polarization information for multi-view consistency evaluation regarding the surface normal.

5. Conclusion

In this paper, we have proposed PolarPMS, which is a new PatchMatch MVS method that can improve the accuracy and completeness of reconstructed 3D models, compared with existing PatchMatch MVS methods. PolarPMS exploits polarization information to improve the estimation accuracy of the object's surface normal and depth, making use of the physical relationship between the observed AoP



Figure 10. Examples of input images and the visual comparison using real data.



PolarPMS (Ours)

Figure 11. Examples of input images and the visual comparison for real car data provided by the authors of Polarimetric MVS.

and the azimuth angle of the normal. Since this relationship is independent of the surface texture, PolarPMS realizes better completeness, especially for texture-less regions. Our PolarPMS has a limitation in reconstructing the regions where there exist strong inter-reflections, which have a more complicated status in polarization, and this encourages us to adopt a more accurate reflection model in our future work.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grant Number 21K17762. The authors would like to thank Dr. Zhaopeng Cui for sharing the data of Polarimetric MVS.

References

- [1] http://graphics.stanford.edu/data/ 3Dscanrep/.5
- [2] Henrik Aanæs, Rasmus Ramsbøl Jensen, George Vogiatzis, Engin Tola, and Anders Bjorholm Dahl. Large-scale data for multiple-view stereopsis. Int. Journal of Computer Vision, 120(2):153-168, 2016. 6
- [3] Sameer Agarwal, Keir Mierle, and The Ceres Solver Team. Ceres Solver, 2023. 3

- [4] Gary A Atkinson and Edwin R Hancock. Shape estimation using polarization and shading from two views. *IEEE Trans.* on Pattern Analysis and Machine Intelligence, 29(11):2001– 2017, 2007. 2
- [5] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. on Graphics (TOG)*, 28(3):24, 2009. 1, 2
- [6] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo – Stereo matching with slanted support windows. In *Proc. of British Machine Vision Conference* (*BMVC*), pages 1–11, 2011. 1, 2
- [7] Xu Cao, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Multi-view azimuth stereo via tangent space consistency. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 825–834, 2023. 2
- [8] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 1558–1567, 2017. 1, 2, 7
- [9] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. PANDORA: Polarization-aided neural decomposition of radiance. In *Proc. of European Conf. on Computer Vision* (ECCV), pages 538–556, 2022. 2
- [10] Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. DeepPruner: Learning efficient stereo matching via differentiable patchmatch. In *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pages 4384–4393, 2019. 2
- [11] Yoshiki Fukao, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Polarimetric normal stereo. In *Proc. of IEEE Conf.* on Computer Vision and Pattern Recognition (CVPR), pages 682–690, 2021. 2
- [12] Silvano Galliani, Katrin Lasinger, and Konrad Schindler. Massively parallel multiview stereopsis by surface normal diffusion. In *Proc. of IEEE Int. Conf. on Computer Vision* (*ICCV*), pages 873–881, 2015. 1, 2
- [13] Jae Yong Lee, Joseph DeGol, Chuhang Zou, and Derek Hoiem. PatchMatch-RL: Deep MVS with pixelwise depth, normal, and visibility. In *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pages 6158–6167, 2021. 2
- [14] Andreas Ley, Ronny Hänsch, and Olaf Hellwich. SyB3R: A realistic synthetic benchmark for 3D reconstruction from images. In *Proc. of European Conf. on Computer Vision* (ECCV), pages 236–251, 2016. 6
- [15] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. ACM Trans. on Graphics (TOG), 38(6):1–17, 2019. 5
- [16] Andrea Romanoni and Matteo Matteucci. TAPA-MVS: Textureless-aware patchmatch multi-view stereo. In *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pages 10413– 10422, 2019. 2
- [17] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 4104–4113, 2016. 2

- [18] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *Proc. of European Conf. on Computer Vision (ECCV)*, pages 501–518, 2016. 1, 2, 3, 4, 5
- [19] William Smith, Ravi Ramamoorthi, and Silvia Tozza. Height-from-polarisation with unknown lighting or albedo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 41(12):2875–2888, 2019. 2
- [20] Chaoran Tian, Weihong Pan, Zimo Wang, Mao Mao, Guofeng Zhang, Hujun Bao, Ping Tan, and Zhaopeng Cui. DPS-Net: Deep polarimetric stereo depth estimation. In *Proc. of IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, pages 3569–3579, 2023. 2
- [21] Fangjinhua Wang, Silvano Galliani, Christoph Vogel, Pablo Speciale, and Marc Pollefeys. PatchmatchNet: Learned multi-view patchmatch stereo. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 14194–14203, 2021. 2
- [22] Jian Wei, Benjamin Resch, and Hendrik PA Lensch. Multiview depth map estimation with cross-view consistency. In *Proc. of British Machine Vision Conference (BMVC)*, pages 1–13, 2014. 1, 2
- [23] Qingshan Xu, Weihang Kong, Wenbing Tao, and Marc Pollefeys. Multi-scale geometric consistency guided and planar prior assisted multi-view stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 45(4):4945–4963, 2022. 2, 6
- [24] Qingshan Xu and Wenbing Tao. Multi-scale geometric consistency guided multi-view stereo. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 5483–5492, 2019. 2
- [25] Qingshan Xu and Wenbing Tao. Planar prior assisted patchmatch multi-view stereo. AAAI Conf. on Artificial Intelligence (AAAI), 34(07):12516–12523, 2020. 2
- [26] Luwei Yang, Feitong Tan, Ao Li, Zhaopeng Cui, Yasutaka Furukawa, and Ping Tan. Polarimetric dense monocular SLAM. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 3857–3866, 2018. 2
- [27] Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric multi-view inverse rendering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 45(7):8798–8812, 2023. 1, 2, 4, 5, 7
- [28] Enliang Zheng, Enrique Dunn, Vladimir Jojic, and Jan-Michael Frahm. PatchMatch based joint view selection and depthmap estimation. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 1510–1517, 2014. 1, 2