Polarimetric Relative Pose Estimation

Zhaopeng Cui[1]   Viktor Larsson[1]   Marc Pollefeys[1,2]
[1]Department of Computer Science, ETH Zurich   [2]Microsoft

Abstract

In this paper we consider the problem of relative pose estimation from two images with per-pixel polarimetric information. Using these additional measurements we derive a simple minimal solver for the essential matrix which only requires two point correspondences. The polarization constraints allow us to pointwise recover the 3D surface normal up to a two-fold ambiguity for the diffuse reflection. Since this ambiguity exists per point, there is a combinatorial explosion of possibilities. However, since our solver only requires two point correspondences, we only need to consider 16 configurations when solving for the relative pose. Once the relative orientation is recovered, we show that it is trivial to resolve the ambiguity for the remaining points. For robustness, we also propose a joint optimization between the relative pose and the refractive index to handle the refractive distortion. In experiments, on both synthetic and real data, we demonstrate that by leveraging the additional information available from polarization cameras, we can improve over classical methods which only rely on the 2D-point locations to estimate the geometry. Finally, we demonstrate the practical applicability of our approach by integrating it into a state-of-the-art global Structure-from-Motion pipeline.

1. Introduction

Estimating the relative pose between two images is a classical problem in computer vision. The epipolar geometry is completely described by the *essential matrix*, which can be minimally estimated from five 2D-point correspondences [21, 25, 10]. Once the essential matrix is recovered, it can be decomposed into the relative rotation and translation [11]. In this paper we propose to exploit the additional geometric information available in polarimetric images for essential matrix estimation.

Different from normal color images, polarimetric images directly encode information of the surface normals for a wide range of materials. In the computer vision literature there have been several papers which show that the encoded normal information can be useful in geometric estimation
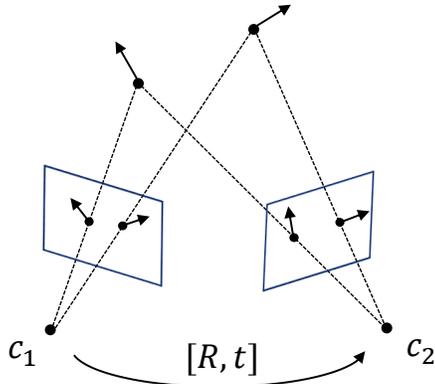
Figure 1. Polarization cameras allow us to recover the 3D surface normals from the 2D images. Using this additional geometric information we show that it is possible to estimate the relative pose using only two 2D point correspondences.

problems. Kadambi *et al*. [13] use the surface normal information from polarization images for depth image enhancement. Cui *et al*. [5] exploit the partial normal information, i.e. phase angles, to propagate depth information along iso-depth contours for multi-view stereo. Yang *et al*. [31] further extend this idea for dense SLAM system. Chen *et al*. [4] study the usage of phase angles for three-view geometry. Different from these papers, we will try to exploit polarimetric information for relative pose estimation between two views.

There are mainly two challenges for using polarimetric information in relative pose estimation. First, extracting the surface normals from the polarization images requires knowledge of the refractive index, also known as refractive distortion [13]. The refractive index depends on the material and is in general unknown, but typically lies in the interval $[1.3, 1.6]$ [1]. To further complicate the problem, this refractive index can vary pointwise in the scene due to the inhomogeneous texture and other material properties.

Secondly, as was described in [13, 5], there is an unavoidable π-ambiguity for each azimuth angle. This azimuthal ambiguity will lead to two possible polarization normals for each point. Additionally, this ambiguity is independent between the two images, even for corresponding points in the two views. This leads to a combinatorial explo-

sion of choices, since for each point correspondence we will have four possible configuration of normals. Thus given n pairs of corresponding point, there will be 4^n possibilities. If the reflection is specular dominated, there will be additional ambiguities in the zenith angle.

In this paper, we will only consider the case for diffuse reflection due to the fact that in practice most points which we are able to match between different viewpoints are diffuse reflection dominated. The key insight in our work is that once the relative rotation is known, resolving the ambiguity is trivial. Conversely, if the true azimuth angles are known, the resulting surface normals greatly simplify the problem of estimating the relative rotation. To address this typical chicken-and-egg problem, we show that it is possible to side-step the combinatorial problem by proposing a method for estimating the relative rotation (and translation) from only two point correspondences. For these two correspondences we only have 16 possible choices of normals for diffuse reflection, making it tractable to evaluate each of them. Once the relative rotation is recovered, we can resolve the azimuthal ambiguity for the remaining point correspondences.

While the refractive index is typically unknown in practice, we show in experiments that the performance of our approach only deteriorates slightly when using incorrect refractive indices. This is consistent with the observation made in [13]. However, to achieve high-accuracy pose estimates we propose to perform joint optimization over both the relative pose and the pointwise refractive indices.

In experiments we show that by using the additional geometric information available in polarization cameras, we can achieve more accurate and more robust estimates of the epipolar geometry compared to the state-of-the-art approach using the 5-point essential matrix solver when the polarimetric measurements are sufficiently good. We evaluate our approach on synthetic data as well as real polarization images. Finally, we show the practical applicability of our approach by integrating it into a state-of-the-art global Structure-from-Motion pipeline [6].

2. Related Work

Relative pose estimation. The standard approach for performing robust estimation of epipolar geometry is using a *hypothesize-and-test* framework, such as RANSAC [8], together with minimal solvers. For essential matrix estimation, the minimal problem requires five 2D-point correspondences. One of the first minimal solvers was presented in [21] by Nistér. Since then there have been many follow-up works, *e.g.*, [25, 10], improving on the original solver.

There are also some methods which try to use additional geometric information besides the epipolar constraints given by 2D-point correspondences. In [23, 26], the authors propose solvers which estimate the relative pose

from three 2D-point correspondences under the assumption of known gravity direction. Their problem formulation is similar to ours in the sense that they also derive additional geometric constraints from knowing the relative orientation of 3D vectors (gravity direction vs. surface normals). However, since they only measure a single 3D vector, this only yields two constraints on the rotation, making the solver require three 2D-point correspondences in total.

The 'quiver'-based motion estimation [12, 15] is also similar to our problem, where n-quiver means a 3D point and n directions from this point. Different from these methods for absolute pose estimation, we utilize the ambiguous surface normal information recovered from polarization images for relative pose estimation.

For relative pose estimation there are also optimal methods which are guaranteed to find the solution with the maximum number of inliers, see *e.g.* [30, 9, 3]. However, these methods are very computationally expensive which limits their usage in practical applications.

Polarimetric 3D modeling. As the polarimetric information encodes 3D surface normal, it has been exploited in several 3D algorithms. Early methods [18, 1, 20] use geometric priors, *e.g.*, the surface normals on the boundary and convexity of the objects, to guide the shape estimation. Later the polarimetric information is combined with shape-from-shading [17, 24] to solve the ambiguity in the surface normal estimation and recover 3D shapes. Some recent works try to utilize polarimetric information in other tasks, *e.g.*, the multiple-view stereo [5], depth enhancement [13], and dense SLAM reconstruction [31]. Most of the methods assume the incident illumination is unpolarized, while the linearly polarized incident light is studied for surface reflectance and normal estimation in [22, 2].

Very recently, Chen *et al*. [4] study the connection between polarization and three-view geometry. The authors consider a weaker geometric constraint by only considering from the polarization phase information. In contrast to [4], we try to exploit the full geometric information available from the polarization cameras and study the relative pose estimation for two views. Note that the phase-angle constraint used in [4] is not applicable in the two-view setting we consider.

3. Preliminaries

As it is shown in [5], if we capture an image through a linear polarizer at a polarization angle ϕ_{pol} under unpolarized illumination, the measured radiance at a single image point is

$$I(\phi_{pol}) = \frac{I_{max} + I_{min}}{2} + \frac{I_{max} - I_{min}}{2} \cos(2(\phi_{pol} - \phi)),$$

(1)

where I_{max} and I_{min} is the maximum and minimum measured radiance, and ϕ is the phase angle. Normally we can

solve for these three unknowns with at least three polarization images with different polarization angles.

3.1. Azimuth Angle Estimation

The azimuth angle φ of surface normal is usually defined as the angle between the projected surface normal direction and x-axis direction in the 2D image. As shown in previous work [13, 5], we can compute the azimuth angle φ as ϕ or $\phi + \pi$ for the diffuse reflection. The π−ambiguity is caused by the factor of 2 within the cosine in Equation 1. For the specular reflection, the azimuth angle is computed as $\varphi \pm \frac{\pi}{2}$. So we can see that no matter for diffuse or specular reflection, there are two candidate values for the azimuth angles which will lead to the ambiguity in the surface normal estimation.

3.2. Zenith Angle Estimation

The zenith angle θ of surface normal is defined as the angle between the surface normal and the negative viewing direction. It is related to the degree of the polarization ρ, which is defined as

$$\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{2}$$

As it is shown in [13, 27], for the diffuse reflection, the zenith angle θ relates to the degree of the polarization ρ as the following,

$$\rho = \frac{(n - \frac{1}{n})^2 \sin^2 \theta}{2 + 2n^2 - (n + \frac{1}{n}) \sin^2 \theta + 4 \cos \theta \sqrt{n^2 - \sin^2 \theta}}, \tag{3}$$

where n is the refractive index. We can compute a single zenith angle once ρ and n are given. Normally the refractive index n for dielectric objects is between 1.3 to 1.6 [1].

For the sepcular reflection, the relationship between θ and ρ is as the following [27],

$$\rho = \frac{2 \sin \theta \tan \theta \sqrt{n^2 - \sin^2 \theta}}{n^2 - 2 \sin^2 \theta + \tan^2 \theta}. \tag{4}$$

There are usually two real solutions for θ given ρ which lead to additional ambiguities in the specular case.

3.3. Surface normal estimation

Suppose zenith angle θ and azimuth angel φ are known, we can compute the surface normal in the local coordinate system of the cameras as [13, 16]:

$$\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} \cos \varphi \sin \theta \\ -\sin \varphi \sin \theta \\ -\cos \theta \end{pmatrix}. \tag{5}$$

From previous sections, we know that there is only a π ambiguity in the azimuth angle for the diffuse reflection. As

a result, there are 2 possible surface normals for the diffuse dominant point. In the contrast, for the specular reflection, there are both ambiguities in the azimuth angle and zenith angle. So there are 4 possible surface normals for the specular dominant point. To use the normals for relative pose estimation we must take these ambiguities into account.

4. Polarimetric Relative Pose Estimation

We now present our method for relative pose estimation exploiting polarimetric information. In Section 4.1 we present a new minimal solver for estimating the relative pose from only two-point correspondences. Next we show how we can resolve the π-ambiguity for the phase angle in Section 4.2. Finally, in Section 4.3 we propose a local refinement scheme for optimizing over the relative pose as well as the pointwise refractive indices. In this section, we only consider the case for the diffuse reflection.

4.1. Relative Pose from Two Point Correspondences

In this section we show how to recover the relative pose from two point correspondences by utilizing the additional geometric constraints from the polarization camera.

For each point correspondence, $(\mathbf{x}, \mathbf{x}')$, we can also compute the 3D surface normals, $(\mathbf{v}, \mathbf{v}')$ in the two views, assuming that the refractive index n is known and the azimuth angle ambiguity is resolved. Since the normals are recovered in the local camera coordinate systems, we get the following constraint on the relative rotation $\mathbf{R} \in SO(3)$,

$$\mathbf{R}\mathbf{v}_i = \mathbf{v}_i', \quad i = 1, 2 \tag{6}$$

Due to noise there will in general not be any exact solution to Equation 6 (even for the case of only two pairs of normals). Instead we consider the following optimization problem,

$$\min_{\mathbf{R} \in SO(3)} \|\mathbf{R}\mathbf{v}_1 - \mathbf{v}_1'\|^2 + \|\mathbf{R}\mathbf{v}_2 - \mathbf{v}_2'\|^2, \tag{7}$$

which seeks to optimally align the two normal pairs. This optimization problem has a closed form solution given in terms of the singular value decomposition (see [7]),

$$\mathbf{U}\Sigma\mathbf{V}^\top = \mathbf{v}_1'\mathbf{v}_1^\top + \mathbf{v}_2'\mathbf{v}_2^\top, \tag{8}$$

then $\mathbf{R} = \mathbf{U} \operatorname{diag}\left(1, 1, \det(\mathbf{U}\mathbf{V}^\top)\right) \mathbf{V}^\top$. Once the rotation is recovered, it is trivial to estimate the translation from two points [14]. Rewriting the epipolar constraints as

$$\mathbf{x}_i' \cdot (\mathbf{t} \times \mathbf{R}\mathbf{x}_i) = \mathbf{t} \cdot (\mathbf{R}\mathbf{x}_i \times \mathbf{x}_i') = 0, \quad i = 1, 2 \tag{9}$$

it becomes clear that the translation can be found as

$$\mathbf{t} = (\mathbf{R}\mathbf{x}_1 \times \mathbf{x}_1') \times (\mathbf{R}\mathbf{x}_2 \times \mathbf{x}_2'). \tag{10}$$

This gives us a single essential matrix assuming the azimuth angle ambiguity is resolved. However, in practice it is typically not the case. Nevertheless, since we only require two point correspondences to estimate the essential matrix, we only need to consider $4^2 = 16$ possible choices. Thus we can estimate an essential matrix for each choice and check which has the largest consensus set among the rest of the points. Another strategy is to only do hypothesis validation for the solution which gets the smallest alignment error in Equation 7.

4.2. Resolving the Azimuth Angle Ambiguity

For each 2D point there is a two-fold ambiguity in the azimuth angle. Furthermore, this ambiguity is independent between the images, meaning that if we have n point correspondences between the images, we have 4^n possible combinations of angle assignments.

However, if we know the relative rotation (even only approximately), we can easily recover the correct azimuth angles (φ, φ') by considering the alignment error,

$$\|\mathtt{R}\mathtt{v}(\varphi) - \mathtt{v}'(\varphi')\|^2. \qquad (11)$$

For each correspondence we only need to check four cases

$$(\phi, \phi'), \ (\phi + \pi, \phi'), \ (\phi, \phi' + \pi) \text{ and } (\phi + \pi, \phi' + \pi), \quad (12)$$

and select the one which minimizes the alignment residual. Note that this can be done in $\mathcal{O}(n)$ time. This assignment is robust to small errors in the rotation since we are only using it to select the best from these four choices.

4.3. Polarimetric Two-View Local Refinement

In this section we propose a local refinement scheme for optimizing jointly over the relative pose and the refractive indices. Using Levenberg-Marquardt [19] we minimize

$$\min_{\mathtt{R}\in SO(3), \mathtt{t}\in\mathbb{S}^2, \{n_i\}} f(\mathtt{R}, \mathtt{t}, \{n_i\}), \qquad (13)$$

where $f(\mathtt{R}, \mathtt{t}, \{n_i\}) =$

$$f_{samp}(\mathtt{R}, \mathtt{t}) + f_{norm}(\mathtt{R}, \{n_i\}) + f_{prior}(\{n_i\}). \quad (14)$$

The loss function consists of three terms which tie together the classical geometric constraints obtained from the 2D point correspondences with the polarimetric normal information. The first term, $f_{samp}(\mathtt{R}, \mathtt{t})$ is the standard squared Sampson loss [11], which tries to enforce the epipolar constraints for the 2D point correspondences. The second term

$$f_{norm}(\mathtt{R}, \{n_i\}) = \gamma_{normal} \sum_{i=1}^{m} \|\mathtt{R}\mathtt{v}_i(n_i) - \mathtt{v}'_i(n_i)\|^2, \ (15)$$

incorporates the normal information. Note that the normals $\mathtt{v}_i(n_i)$ depend on the refractive indices n_i as shown

in Equation 5. The parameter γ_{normal} is set to 10^{-3} and 10^{-4} for our synthetic and real experiments respectively.

Finally, to make the optimization more robust to poor initialization we add a term which penalizes refractive indices which deviate far from the expected values (most dielectric materials have $n \in [1.3, 1.6]$),

$$f_{prior}(\{n_i\}) = \gamma_{prior} \sum_{i=1}^{m} \left(n_i - n_i^0\right)^2, \qquad (16)$$

where γ_{prior} is the weighting parameter which is set to 10^{-5} and 10^{-4} for our synthetic and real experiments respectively. We set n_i^0 to be 1.5 in our experiments.

4.3.1 Implementation Details

For stopping criterion we use $\|\nabla f\|_\infty < 10^{-8}$ as well as a threshold on the maximum number of iterations (set to 100 in our experiment). To make the cost robust to outlier measurements we use the truncated ℓ_2 loss,

$$\varrho(x) = \begin{cases} x^2 & \text{if } x < \epsilon, \\ \epsilon^2 & \text{otherwise.} \end{cases} \qquad (17)$$

Additionally, in the optimization we consider points where the refractive index $n \notin [1, 2]$ as outliers. For the rotation we use the local angle-axis parameterization,

$$\mathtt{R}(\mathtt{r}) = \mathtt{R}_0 \exp([\mathtt{r}]_x), \qquad (18)$$

where $\mathtt{R}_0$ is the rotation from the previous iteration. Similarly, to avoid the scale-freedom in the translation, in each iteration we parameterize the translation locally using the tangent space of the unit sphere,

$$\mathtt{t}(x, y) = \mathtt{t}_0 + \mathtt{b}_1 x + \mathtt{b}_2 y, \qquad (19)$$

where $\mathtt{b}_1$ and $\mathtt{b}_2$ are two unit vectors orthogonal to $\mathtt{t}_0$.

If we use independent refractive indices n_i for each point there will be $3 + 2 + N_p$ unknowns in our optimization problem, where N_p is the number of 2D correspondences. In the case where the refractive index is constant over all points, we would only have 6 unknowns.

As is the case for standard Bundle Adjustment [28], we can employ the Shur complement trick to greatly reduce the computational cost of the iterations. In our case the linear update equations in LM are given by

$$\begin{bmatrix} \mathtt{J}_{\mathtt{Rt}}^T \mathtt{J}_{\mathtt{Rt}} + \lambda \mathtt{I} & \mathtt{J}_{\mathtt{Rt}}^T \mathtt{J}_{\mathtt{n}} \\ \mathtt{J}_{\mathtt{n}}^T \mathtt{J}_{\mathtt{Rt}} & \mathtt{J}_{\mathtt{n}}^T \mathtt{J}_{\mathtt{n}} + \lambda \mathtt{I} \end{bmatrix} \begin{bmatrix} \Delta_{\mathtt{Rt}} \\ \Delta_{\mathtt{n}} \end{bmatrix} = \begin{bmatrix} -\mathtt{J}_{\mathtt{Rt}}^T \epsilon \\ -\mathtt{J}_{\mathtt{n}}^T \epsilon \end{bmatrix}, \quad (20)$$

where $\epsilon \in \mathbb{R}^{N_{res}}$ is the vector of residuals, $\mathtt{J}_{\mathtt{Rt}} \in \mathbb{R}^{N_{res}\times 5}$, $\mathtt{J}_{\mathtt{n}} \in \mathbb{R}^{N_{res}\times N_p}$ are the Jacobians w.r.t. the rotation/translation and the refractive indices respectively. Eliminating $\Delta_{\mathtt{n}}$ from the second row we get

$$\Delta_{\mathtt{n}} = -\left(\mathtt{J}_{\mathtt{n}}^T \mathtt{J}_{\mathtt{n}} + \lambda \mathtt{I}\right)^{-1} \left(\mathtt{J}_{\mathtt{n}}^T \epsilon + \mathtt{J}_{\mathtt{n}}^T \mathtt{J}_{\mathtt{Rt}} \Delta_{\mathtt{Rt}}\right). \qquad (21)$$
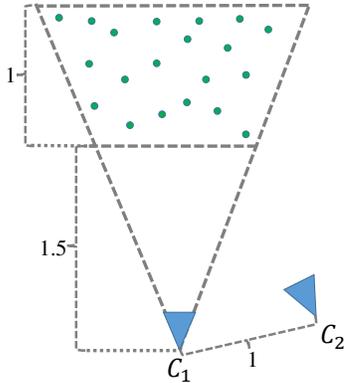
Figure 2. Illustration of the geometric configuration for our synthetic experiment.

Figure 3. Angular rotation error distribution for the test on the synthetic data. The error is measured in degrees.

Note that $H = \left(J_n^T J_n + \lambda I\right)$ is a diagonal matrix since there are no direct interactions between the different refractive indices. So while $H \in \mathbb{R}^{N_p \times N_p}$ will be a large matrix if we have a lot of point correspondences, computing the inverse is extremely cheap (in regular bundle adjustment H will be block-diagonal with 3×3-blocks). Completing the Shur complement, inserting (21) into (20), we get

$$\left(J_{Rt}^T J_{Rt} + \lambda I - J_{Rt}^T J_n H^{-1} J_n^T J_{Rt}\right) \Delta_{Rt}$$
$$= -(J_{Rt}^T - J_{Rt}^T J_n H^{-1} J_n^T)\epsilon, \qquad (22)$$

which then reduces the problem to solving a 5×5 linear system, regardless of the number of points N_p.

After each iteration in Levenberg-Marquardt we update the angle ambiguity using the current rotation estimate as described in Section 4.2. Note that this only affects the f_{norm} term, and will always lead to an equal or lower function value for f.

5. Experiment

5.1. Evaluation on Synthetic Data

We first quantitatively evaluate our method with two synthetic images with known ground truth poses. The test geometry is shown in Figure 2. The camera coordinate of the first camera is set to be the same as the world coordinate. The second camera is placed at a random location around the first camera, and the baseline length between the two cameras are set to 1 unit. The maximum angle along x, y and z axes are set to be $30°$, $40°$ and $5°$ respectively. We randomly generate the scene points within the viewing volume of the first camera with a depth range from 1.5 to 2.5 unit. The synthetic image resolution is 352×288 pixels and the field of view is set to be $45°$. We get image coordinates of each 3D point through projection. Moreover, we also randomly generate a normal vector facing towards the first camera for each point. From the normal vectors, we
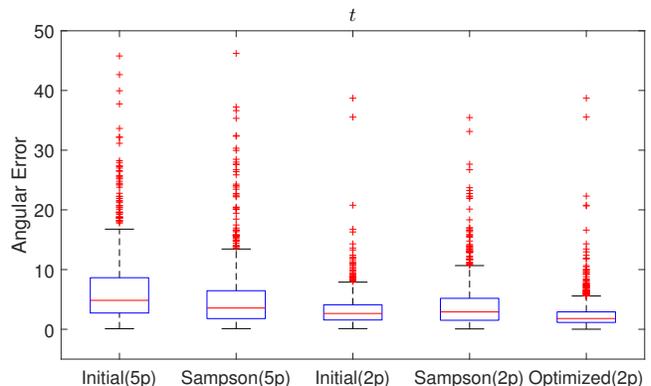
Figure 4. Angular translation error distribution for the test on the synthetic data. The error is measured in degrees.

can compute the phase angle and degree of polarization for each point in two cameras. We randomly select a refractive index between 1.3 and 1.7.

We assume that all points are diffuse-dominant and that they share the same refractive index which is unknown beforehand. We use 1.5 as our initial guess for the refractive index. Moreover, we add zero mean Gaussian noise to image coordinates and phase angles. The standard deviation for image coordinates is set to be 2 pixels, and the standard deviation of phase angles is set to be $3°$. For degrees of polarization, we add multiplicative noise with the standard deviation to be 5%.

We evaluate the accuracy of the estimated poses by computing the angular error to the ground truth for both the rotation and translation. We conduct 1000 random trials for the test.

Comparison with 5-point algorithm [21]. We first compare our algorithm with the efficient 5-point algorithm [21]. The performance of both methods are shown in Figure 3 and Figure 4. The mean errors are shown in Table 1. We report the errors after the initial RANSAC, after optimization on the Sampson error, and finally with our optimiza-

	5-point		2-point		
	Initial	Sampson	Initial	Sampson	Optimized
R_{err}	6.10	4.95	2.30	3.59	**1.80**
t_{err}	9.30	7.37	3.25	4.08	**2.52**

Table 1. Comparison with 5-point algorithm on the synthetic data. The error is measured in degrees.

(a) Angular rotation error (in degrees)　　　　(b) Angular translation error (in degrees)
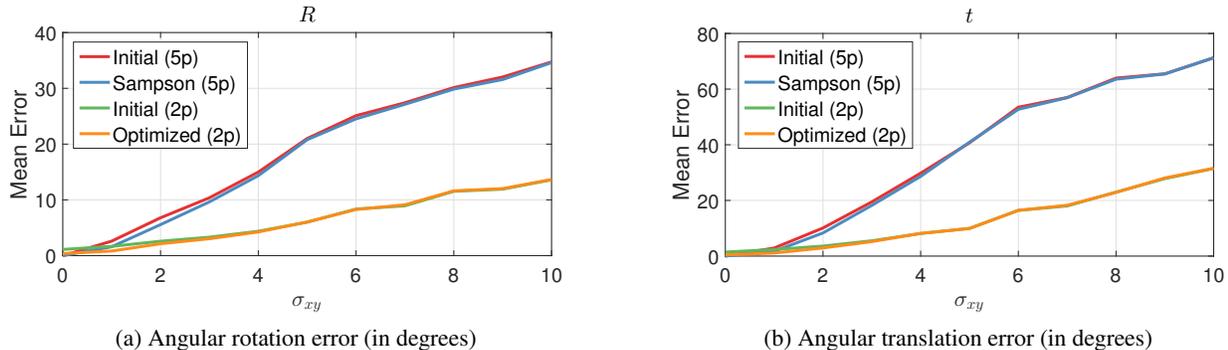
Figure 5. Performance with different levels of noise in image coordinates.

tion scheme from Section 4.3 which also promotes geometric consistency of the normals. We can see that our algorithm consistently outperforms the 5-point algorithm in this experiment setup, with and without the non-linear refinement. We can also see that only optimizing the Sampson error becomes worse compared to using the proposed optimization. While the geometric information from polarization only gives constraints for the relative orientation, we can see that it also benefits the relative translation estimation.

We also test the robustness of the different methods under varying levels of noise in the point correspondences. Figure 5 shows the average angular errors as we vary the standard deviation of the noise level from 0 to 10 px. From the figure, we can see that with reasonable measurements of polarization information, our 2-point algorithm is more robust against noise in the image coordinates.

Finally, we compare the average numbers of RANSAC iterations and runtimes of different methods. When the desired probability is set to 0.99, the average numbers of RANSAC iterations for 5-point algorithm and ours are 17.8 and 7.8 respectively. The mean runtimes without optimization of 5-pt algorithm and ours are 7.1 ms and 14.3 ms on a standard desktop PC in unoptimized MATLAB code. We believe our method could be improved significantly by using a closed form solution for the two-vector alignment problem instead of explicitly computing the SVD.

Robustness against noise in different polarization measurements. Since our 2-point algorithm relies on extra polarization measurements, we also test the robustness of our algorithm under different noise levels of the polarization measurements. We first test the robustness of our algorithm

against noise in the phase angle. We change the standard deviation of noise for the phase angle from 0 to 15 degrees, and the performance is shown in Figure 6. We can see that our algorithm performs robustly with noise in the phase angle measurement. Even with a standard deviation of $15°$, the errors in the relative pose estimation is still rather small, with a mean error of $4.13°$ and $5.08°$ in rotation and translation after optimization.

Next we test the robustness against noise in the measurement of degree of polarization. We change the standard deviation from 0% to 15%, and the performance is shown in Figure 7. We can see that our algorithm is also quite robust against the noise in the degree of polarization.

Performance with different initial refractive indices. Since the refractive index is generally unknown, it is particularly important that the method could gracefully handle incorrect refractive indices. We also studied the performance of our algorithm with different initial refractive index. In order to have a clear understanding of the impact of the initial refractive index, we assume all the other measurements are noiseless, and set the ground-truth refractive index to be 1.5. Then we change the initial guess of refractive index from 1.3 to 1.7 with a step size of 0.02. The results are shown in Figure 8. From Figure 8, we can see that with a correct guess of refractive index, the errors in both initial translation estimation and initial rotation estimation are close to zero. When the initial guess of refractive index is biased from the ground truth, the initial pose estimation from the 2-point algorithm becomes worse, but still maintains relative small errors and these errors could be handled with our polarimetric two-view optimization. From this experiment, we can see that the incorrect guess of the refractive index
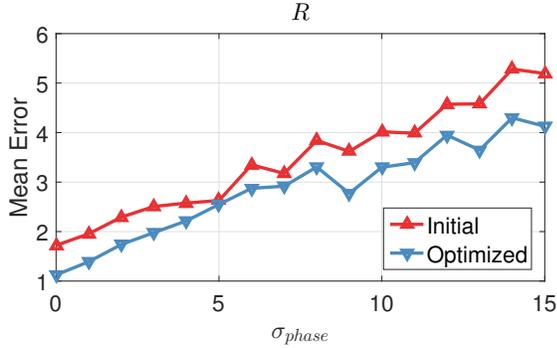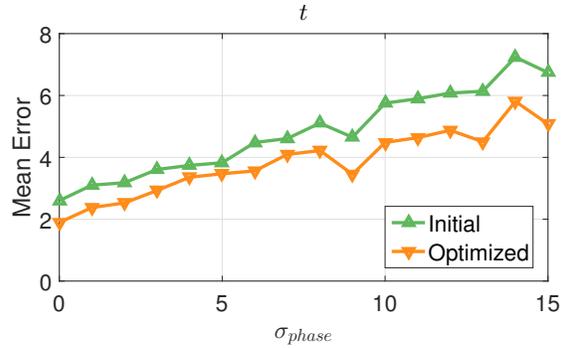
(a) Angular rotation error (in degrees) (b) Angular translation error (in degrees)

Figure 6. Performance with different levels of noise in phase angles.
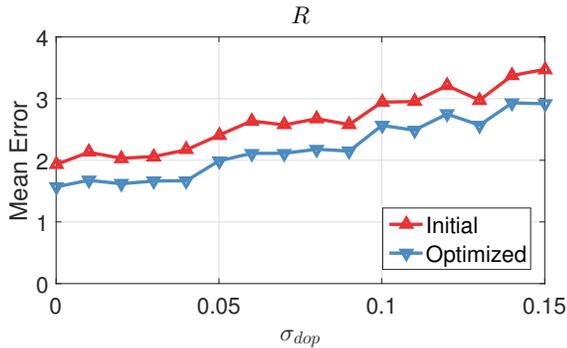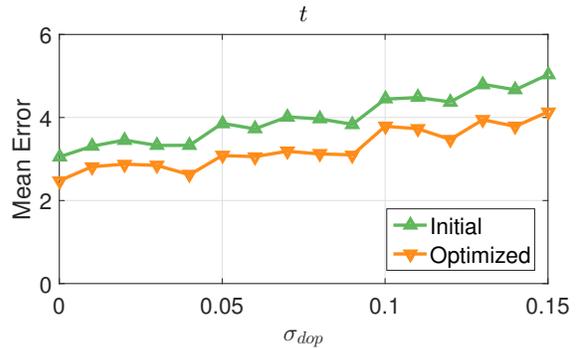
(a) Angular rotation error (in degrees) (b) Angular translation error (in degrees)

Figure 7. Performance with different levels of noise in degree of polarization.

only influences the initial pose estimation to some extent, and this could be solved by the later joint optimization in the proposed method. Our observation is similar to that in [13]. In [13], the authors shows that the difference between the ground-truth refractive index and hard-coded refractive index exhibits a certain distortion in the surface normal estimation. We also compute the mean error for the estimated refractive index, and it is 8.27% after the optimization with an initial error of 9.09%.

Performance with non-uniform refractive index. Next we consider the especially challenging case when all the points have their own refractive index. In this case, there are more parameters to be optimized in our joint optimization. With the same setup as the initial synthetic experiment for all the other measurements, we generate different refractive indices for each point in the interval [1.4,1.6]. The average angular errors for rotation and translation are $3.66°$ and $4.91°$. Although it is worse than the case with uniform refractive index, the performance is still better than the 5-point algorithm.

5.2. Evaluation on Real Data

We further evaluate our method with real polarization datasets from [5], which were captured by a Cannon EOS 7D camera with a $50mm$ lens and a Hoya linear polarizer

in the front of the camera lens. VASE has 36 images, BAL-LOON has 24 images, and TALLVASE has 36 images. These three datasets are dielectric and relatively larger compared to the other datasets in [5]. For each viewpoint, there are seven images captured with the polarization angle spaced $30°$ degrees apart, from which we can compute the phase angle and degree of polarization.

The datasets also provide global camera poses recovered by the incremental structure-from-motion method from [29]. Since these have been optimized in bundle adjustment with constraints from multiple views, we take the relative poses from this reconstruction as the ground truth for our experiment. Moreover, for a reasonable comparison, we only consider image pairs with sufficiently large shared field of view. In total we have 146 pairs of images for the VASE dataset, 57 pairs for BALLOON, and 124 pairs for TALL-VASE. In Figure 9 we show some example images from the three datasets.

The results of the proposed method and the 5-point algorithm are shown in Table 2. From the table, we can see that our method outperforms the 5-point algorithm in terms of both translation and rotation angular errors. Especially for the TALLVASE dataset, our method reduces the rotation error by 44.39%, and the translation error by 24.94%. We believe this is mainly because the dataset has significant
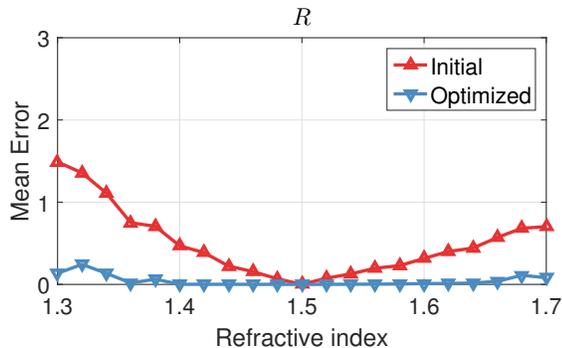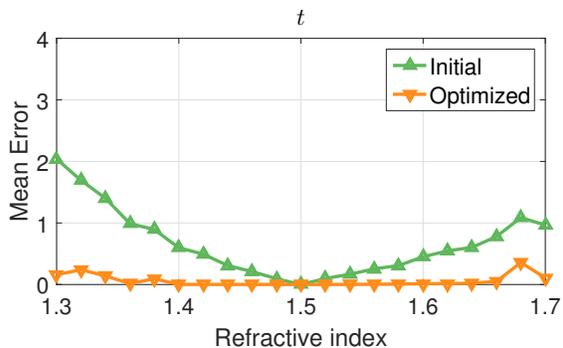
(a) Angular rotation error (in degrees)   (b) Angular translation error (in degrees)

Figure 8. Performance with different initial guess of the refractive index.

Figure 9. Sample images and dense reconstruction results of the real datasets.

| | 5-point | | 2-point | |
	R_{err}	t_{err}	R_{err}	t_{err}
VASE	14.70	21.95	**14.44**	**21.23**
BALLOON	7.44	7.51	**6.87**	**6.85**
TALLVASE	22.91	17.36	**12.74**	**13.03**

Table 2. Accuracy comparison on the real datasets.

repetitive features and higher outlier ratios which favors the smaller sample size of our solver compared to the 5-point algorithm.

In order to further test the computed relative poses, we feed the computed relative poses into a global structure-from-motion pipeline [6]. We measure the error of the re-covered global poses compared to the given global poses.

The mean angular error for the global rotation is $0.97°$, $0.81°$ and $2.57°$ for VASE, BALLOON and TALLVASE respectively. The mean error for the global camera position is 0.010, 0.009 and 0.028 if we set the maximum distance of two cameras to be 1. We can see that both global rotation and camera position errors are quite small, which validates the effectiveness of the proposed 2-point algorithm for the relative pose estimation.

We also conduct the dense reconstruction by integrating both the geometric and polarimetric information as [5]. The global poses recovered from [6] initialized with the epipolar geometries estimated using the proposed approach are used. The final reconstructed dense models are shown in Figure 9, and we can see that the dense 3D models can be accurately recovered.

6. Conclusion

In this paper, we propose a minimal solver for the relative pose estimation from polarimetric images. By exploiting the encoded surface normal information in the polarization images, we can recover the relative pose given two point correspondences. With the requirement of only two point correspondences, our algorithm could easily solve the ambiguity in the surface normal estimation. Moreover, in order to handle the refractive distortion, we also propose a joint optimization framework between the relative pose and the refractive index. We validate our algorithm with both synthetic and real datasets. Experiments show that our algorithm is very robust and performs better than the 5-point algorithm when the polarimetric measurements are sufficiently good. We only studied the case for the diffuse reflection in this paper; however our algorithm can be easily adapted for the specular reflection, and we will study this in the future.

References

[1] Gary A. Atkinson and Edwin R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE transactions on image processing*, 15(6):1653–1664, 2006.

[2] Seung-Hwan Baek, Daniel S Jeon, Xin Tong, and Min H Kim. Simultaneous acquisition of polarimetric svbrdf and normals. *ACM Transactions on Graphics (TOG)*, 37(6):268-1, 2018.

[3] Jesus Briales, Laurent Kneip, and Javier Gonzalez-Jimenez. A certifiably globally optimal solution to the non-minimal relative pose problem. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2018.

[4] Lixiong Chen, Yinqiang Zheng, Art Subpa-asa, and Imari Sato. Polarimetric three-view geometry. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 20–36, 2018.

[5] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 1558–1567, 2017.

[6] Zhaopeng Cui and Ping Tan. Global structure-from-motion by similarity averaging. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 864–872, 2015.

[7] David W Eggert, Adele Lorusso, and Robert B Fisher. Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine vision and applications*, 9(5-6):272–290, 1997.

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

[9] Johan Fredriksson, Viktor Larsson, Carl Olsson, and Fredrik Kahl. Optimal relative pose with unknown correspondences. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] Richard Hartley and Hongdong Li. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 34(12):2303–2314, 2012.

[11] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[12] Björn Johansson, Magnus Oskarsson, and Kalle Åström. Structure and motion estimation from complex features in three views. In *Proc. of ICVGIP*, 2002.

[13] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3D: High-quality depth sensing with polarization cues. In *Proc. of International Conference on Computer Vision (ICCV)*, 2015.

[14] Laurent Kneip, Margarita Chli, and Roland Y Siegwart. Robust real-time visual odometry with a single camera and an imu. In *Proc. of the British Machine Vision Conference*. British Machine Vision Association, 2011.

[15] Yubin Kuang and Kalle Åström. Pose estimation with unknown focal length using points, directions and lines. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 529–536, 2013.

[16] Fotios Logothetis, Roberto Mecca, Fiorella Sgallari, and Roberto Cipolla. A differential approach to shape from polarisation: A level-set characterisation. *International Journal of Computer Vision (IJCV)*, pages 1–14, 2018.

[17] Ali H. Mahmoud, Moumen T. El-Melegy, and Aly A. Farag. Direct method for shape recovery from polarization and shading. In *Proc. of International Conference on Image Processing*, 2012.

[18] Daisuke Miyazaki, Robby T. Tan, Kenji Hara, and Katsushi Ikeuchi. Polarization-based inverse rendering from a single view. In *Proc. of International Conference on Computer Vision (ICCV)*, 2003.

[19] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. *Numerical analysis*, pages 105–116, 1978.

[20] Olivier Morel, Fabrice Meriaudeau, Christophe Stolz, and Patrick Gorria. Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In *Proc. of Machine Vision Applications in Industrial Inspection XIII*, volume 5679, pages 178–186, 2005.

[21] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 26(6):0756–777, 2004.

[22] Jérémy Riviere, Ilya Reshetouski, Luka Filipi, and Abhijeet Ghosh. Polarization imaging reflectometry in the wild. *ACM Transactions on Graphics (TOG)*, 36(6):206, 2017.

[23] Olivier Saurer, Pascal Vasseur, Cedric Demonceaux, and Friedrich Fraundorfer. A homography formulation to the 3pt plus a common direction relative pose problem. In *Proc. of Asian Conference on Computer Vision (ACCV)*, pages 288–301. Springer, 2014.

[24] William A. P. Smith, Ravi Ramamoorthi, and Silvia Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *Proc. of European Conference on Computer Vision (ECCV)*, 2016.

[25] Henrik Stewenius, Christopher Engels, and David Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006.

[26] Chris Sweeney, John Flynn, and Matthew Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In *Proc. of International Conference on 3D Vision (3DV)*, volume 1, pages 483–490. IEEE, 2014.

[27] Vage Taamazyan, Achuta Kadambi, and Ramesh Raskar. Shape from mixed polarization. *arXiv preprint arXiv:1605.02066*, 2016.

[28] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *Proc. of International workshop on vision algorithms*, pages 298–372. Springer, 1999.

[29] Changchang Wu. Towards linear-time incremental structure from motion. In *Proc. of International Conference on 3D Vision*, 2013.

[30] Jiaolong Yang, Hongdong Li, and Yunde Jia. Optimal essential matrix estimation via inlier-set maximization. In *Proc. of European Conference on Computer Vision (ECCV)*, 2014.

[31] Luwei Yang, Feitong Tan, Ao Li, Zhaopeng Cui, Yasutaka Furukawa, and Ping Tan. Polarimetric dense monocular slam. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 3857–3866, 2018.