Depth from Stereo and Polarization for Opaque Non-Lambertian Surfaces

Kai Berger[1], Randolph Voorhies[2] and Larry H. Matthies[1]

***Abstract*— This paper presents a new approach to depth from stereo in scenes with non-Lambertian surfaces. While these surfaces usually prove to be hard to incorporate in photo-consistency energy minimization, we retrieve additional information about their surface normals from applying a polarization-based approach that computes the degree and angle of polarization at each pixel. Therefore, we set up a stereo camera rig that captures four different rotations of a linear polarizer simultaneously for each camera. While most existing polarization-based approaches to shape reconstruction require a setup with known lighting, our approach uses arbitrary unpolarized illumination. We describe a regularization-based stereo approach that constrains specular surface normals in the scene to lie in the plane of their observed angle of polarization. We derive the mathematical foundations of this approach and present results on virtually generated stereo and real-world stereo imagery.**

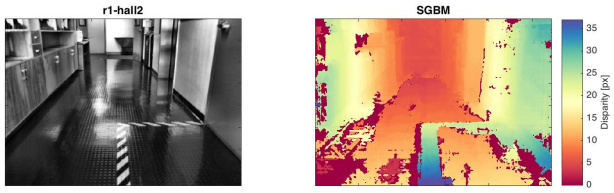

Fig. 1. Specular surfaces (left) usually pose a challenge to state-of-the-art stereo vision algorithms (right, an exemplary result from SGBM [1]): In this case the floor is reconstructed as being further from the camera as they are.

I. Introduction

Specular and transparent surfaces are important classes of materials with non-Lambertian reflectance that are commonplace in urban areas. Many robotics applications will operate in urban areas, so it is important to develop robotic perception systems that function well in the presence of such materials. Passive 3-D perception systems that do not explicitly allow for the presence of such materials perform poorly in these environments; active ranging systems also have difficulty (Fig. 1).

This paper focuses primarily on indoor/outdoor urban scenes with opaque surfaces that include areas of both Lambertian and non-Lambertian reflectance. Several techniques have been explored to improve depth perception in such environments, including multi baseline stereo, active illumination, and polarization. Multi baseline methods have generally sought to overcome the effects of strong highlights by searching for images where corresponding pixels do not contain the highlight. Such techniques require many cameras or many views, which is problematic for robotic vehicles in dynamic environments. Methods that focus on strong highlights are not well-suited to scenes where large areas combine diffuse and specular reflection, such as waxed linoleum floors and analogous textured materials with a surface sheen. For surfaces with both diffuse and specular components of reflection, various forms of active illumination can be used to obtain range measurements from the diffuse reflection, including structured light and lidar. Structured light is a relatively short-range technique, especially wherever ambient illumination is strong, like outdoors; scanning lidar overcomes strong ambient lighting, but still has difficulty with specular surfaces, especially at high angles of incidence. Polarization has been used for depth perception, but so far either as a monocular camera system, where absolute depth perception is still not possible, or as an adjunct to active depth sensors like the Kinect, where it has been used to help correct errors. Compact cameras are now available with microgrid polarization filter arrays on the focal plane that provide four orientations of linear polarization in each 2×2 block of pixels. These can be used to estimate the degree of polarization and the angle of polarization (AOP) in the image for each 2×2 pixel block. The AOP is the image plane projection of the surface normal; therefore, it provides a constraint on the surface normal at each pixel where the AOP is well-defined. We present the first stereo algorithm to use such information. We formulate an objective function that includes a data term, the surface normal constraint from polarization, and a second-order surface smoothness prior, which yields better results in typical man-made environments than first order smoothness prior.

Sections II and III, respectively, discuss related work and the shortcomings of conventional depth from stereo in more detail. We describe how to generate virtual stereo imagery with a polarization-aware ray tracer in section IV. In doing so we are able to have rectified stereo images and ground truth disparities for later comparison at hand. Then we describe the formulation of our regularization-based approach in section V. In section VI we describe the camera setup that we have used to create real-world scenes and evaluate various state-of-the art stereo vision algorithms and our approach on both the virtual and real-world scenes before we conclude in section VII.

[1]Kai Berger and Larry Matthies are with Jet Propulsion Laboratory, California Institute of Technology 4800 Oak Groove Drive Pasadena {kuberger,lhm}@jpl.nasa.gov

[2]Randolph Voorhies is with XXX xxx

II. Related Work

Early work on stereo vision for non-Lambertian scenes focused on strong specular highlights, analyzing how such points may provide information about local surface curvature [2], [3]. Later, trinocular camera configurations were used to attempt to ensure that at least one pair of cameras would image each scene point without being affected by specular highlights [4]; this concept has been extended to more general multibaseline camera configurations [5], [6], [7], [8]. This line of work often considered specularities as sparsely occurring highlights; this does not address the important case where reflections are superimposed over large areas of shiny surfaces that also have intrinsic patterns in the diffuse component of reflection. This was addressed with multibaseline stereo in an approach that assumed the presence of multiple, additive layers and used a nested plane sweep algorithm to estimate the depth of each layer [9]. Results were shown on real data in tabletop-like settings; however, there was no discussion of computational complexity or the number of images required to obtain satisfactory results. In general, for mobile robot applications in dynamic scenes, it is desirable to solve this problem with the smallest number of cameras necessary.

In the context of a single moving camera, specular and diffuse optical flow from the same shiny surface has been estimated as layered flow to recover shape under some simplifying assumptions, including known camera motion, distant illumination, and a twice-differentiable surface [10]. This was demonstrated in simulation and with real images of a shiny ball. We are concerned with relatively arbitrary, discontinuous urban scenes. In principle, stereo algorithms designed to handle transparent or semi-transparent surfaces [11] could be applicable to specular scenes, alone or combined with a layered approach to optical flow; we have not seen this done.

Time of flight laser rangefinders can estimate range to shiny materials that also have a diffuse component of reflection; however, this capability degrades with increasing angle of incidence of the beam on the surface [12] and with increasing degree of specularity. Various methods have been examined to use other forms of structured illumination to estimate range to surfaces with both specular and diffuse reflection [13], [14], [15], [16]. These methods can be effective, but are limited by the range of the illuminator, especially in scenes with strong ambient lighting, particularly outdoors.

Polarization has been used for several purposes in computer vision, including classifying different types of materials and different types of intensity edges, separating specular and diffuse components of reflection, and constraining surface orientation [17], [18]. The angle of polarization (AOP) in an image constrains the surface normal of the corresponding point in space to lie in the specular plane containing the image AOP vector and the center of projection; if the index of refraction of the material is also known, constraints on the full surface normal can also be obtained [17]. These constraints have been used to integrate image-centric height models [19], [20], [21]; however, absolute depth cannot be inferred this way. Related ideas, without knowledge of the index of refraction of the material, were applied to multiview polarization imagery in [22] to reconstruct shape; however, the method was only developed for and applied to textureless objects. Polarization and multiple known point light sources are used in [23] to recover the shape of smooth specular objects. Some recent papers have used polarization-based surface orientation information to improve depth data from the Microsoft Kinect [16], [24]; these methods are also limited by the range of the illuminator in the depth sensor.

The idea of using polarization-based surface normal constraints together with stereo vision was discussed in [17], but no algorithm for estimating depth maps in this way was described; to our knowledge, we are the first to do so. We formulate an energy function that combines a robust data term, a surface smoothness prior, and the constraint on the specular plane available from the AOP measured at each pixel. For urban scenes, second-order smoothness priors generally give better results than first-order priors. Both the surface normal constraint and the second-order smoothness prior lead to ternary terms in the energy function and a difficult optimization problem. We adapt the approach in [25] to rewrite these as combinations of binary terms and we use their estimation algorithm, which is based on quadratic pseudo-boolean optimization, to approximate the optimal depth map. Tests with stereo polarization imagery of a simulated hallway scene and several real scenes show that this yields better results than methods that do not exploit polarization information.

III. Shortcomings of conventional Depth from Stereo

Depth from stereo can be computed by solving an energy minimization problem on the photoconsistency between the left and right camera's image in a stereo rig. In accordance with [25], this can be expressed as

$$min_{D(x)} E_{\text{photo+smoothness}} = \sum_{x} f(I_1^{\Pi}(x, D(x)) - I_0(x), V) + \sum_{\mathcal{N} \in N} W_1(\mathcal{N}) \rho(\mathcal{S}(\mathcal{N}, \mathcal{D})) \tag{1}$$

with $D(x)$ the depth/disparity map, Π the projection matrix between the two cameras and I_0 the image of the left camera. Further, V is a visibility flag and f is a photo-consistency metric (both discussed in greater detail in [25]). $\mathcal{N}(x)$ is the clique containing pixel x, and W_1 and ρ are weighting functions for the smoothness prior $\mathcal{S}$. While the photo consistency formulation holds for diffuse surfaces, the non-Lambertian surfaces will provide intensity values

$$I_0(x_{\text{nonLambetian}}) = \gamma I_{\text{diffuse}} + (1 - \gamma) I_{\text{specular}}. \tag{2}$$

Depending on γ, a $min_{D(x_{\text{nonLambetian}})}$ may be evaluated at a virtual distance to the mirrored object behind the non-Lambertian surface or even to an unpredictable distance.

Fig. 2. The light transport through a synthetic hallway scene generated by our ray tracer. In this scene, all surfaces are purely diffuse except for the floor. A virtual linear polarizer at (from left to right) rotations $0°, 45°, 90°$ and $135°$ is simulated in front of the camera. Through the four images it can be seen, that the intensity of the floor pixels follows a square cosine behavior, and that the reflected highlights are most visible at the $0°$ polarizer rotation.

Reasoning that for most specular surfaces a Fresnel behavior can be assumed, the term I_{specular} should be at least partially polarized when imaged by the camera. Thus, adding a linear polarizer in front of the camera and rotating it perpendicular to the plane of incidence on the non-Lambertian surface will mitigate the specularity. However, the plane of incidence and thus the rotation angle α are unknown. Also, except for Brewster's angle, the reflected light is not exclusively perpendicularly polarized, and thus cannot simply be extinguished by rotating a linear polarizer accordingly.

IV. Providing Ground Truth by Polarized Light Transport Simulation

We seek to test our algorithms on datasets with accurate ground truth, thus we developed a ray tracing solution that is capable of simulating the light transport of polarized light. Our ray tracer incorporates the Fresnel terms for parallel and perpendicular reflection of shiny surfaces and simulates the presence of a linear polarizer in front of the virtual camera. We implement the Fresnel Terms in the well-known form [26] for single-surface media:

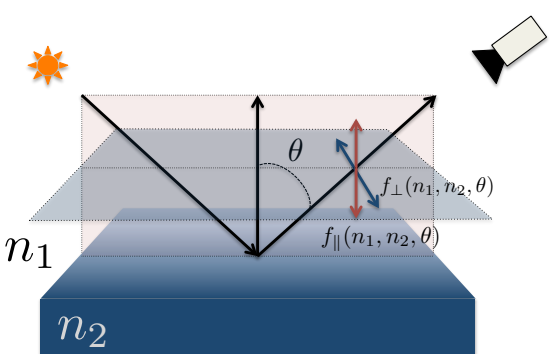

Fig. 3. The we simulate the polarization effects of shiny surfaces by incorporating the Fresnel terms for singe-surface media, Eqns. 3- 5. The intensity of parallel $f_\parallel$ (red) and perpendicular $f_\perp$ (blue) polarized light is dependent on the zenith angle θ for a given surface point. The refractive index of the surface material and surrounding medium are n_2 and n_1.

$$f_\perp(n_1, n_2, \theta) = \frac{\sin^2(\theta - \theta')}{\sin^2(\theta + \theta')} \quad (3)$$

$$f_\parallel(n_1, n_2, \theta) = \frac{\tan^2(\theta - \theta')}{\tan^2(\theta + \theta')} \quad (4)$$

with

$$\theta' = \sin^{-1}\left(\frac{n_1 \sin(\theta)}{n_2}\right) \quad (5)$$

Where θ denotes the zenith angle at a surface point computed between the viewing direction and the surface normal and n_1 and n_2 denote the refractive indices of the surrounding medium and the surface material. The rotation of the azimuth angle α at a surface point needs to be taken into account for the parallel and perpendicular reflection by applying the corresponding rotational change of basis. This way, the ray tracer is capable to accurately simulate the square cosine behavior of the captured intensity for shiny surfaces at arbitrary rotations of a linear polarizer. Since our real-world camera setup provides us with images captured at four equidistant polarizer rotations only, we confine ourselves to generating imagery for virtual polarizer rotations at $0°, 45°, 90°$ and $135°$, Figure 2. The ray tracer performs two tracing steps: in the first step it simulates purely diffuse and unpolarizing surfaces and their contribution to the scene light transport, in the second step the shiny surface and their contribution to the polarization state of the light are simulated. Both results are combined before applying the linear polarizer in front of the virtual camera.

V. Definition of Energy Functional for Graph Cut Solution

In our approach we distinguish two areas in the image: specular areas, where we can see noticable intensity differences for varying rotations of a linear polarizer, and diffuse areas with negligible intensity differences. In the diffuse areas a stereo vision approach like SGBM [1] performs well. In the specular areas, it is likely to produce erroneous results, so we incorporate the Angle of Polarization computed from the four polarizer rotations in these areas. A distinction between the specular and diffuse areas can be performed by thresholding the Degree of Linear Polarization, which is computed from the four polarizer rotations as well.

We define the energy functional as a sum of a photo consistency term and a smoothness in accordance to [25]:

$$\begin{aligned} E_{\text{photo+smoothness}} &= \sum_x f(I_1^{\Pi}(x, D(x)) - I_0(x), V) \\ &+ \sum_{\mathcal{N} \in N} W_1(\mathcal{N}) \rho(\mathcal{S}(\mathcal{N}, \mathcal{D})) \\ &+ \sum_x W_2(\mathcal{N}(x)) \rho(\mathcal{A}(\mathcal{N}(x), \mathcal{D}, \alpha(x))) \end{aligned} \tag{6}$$

with $\alpha(x)$ the angle of polarization at a given pixel x - corresponding to the azimuth α described in Sect. IV, $\mathcal{N}(x)$ the clique containing pixel x and $\mathcal{A}$ defined as

$$\begin{aligned} \mathcal{A}(\mathcal{N}, \mathcal{D}, \alpha) &= tan^{-1}(\Gamma((\Pi^{-1}(p, D(p)) - \Pi^{-1}(q, D(q)) \\ &\times \Pi^{-1}(r, D(r)) - \Pi^{-1}(q, D(q))))) - \alpha), \\ &\{p, q, r\} \in \mathcal{N} \end{aligned} \tag{7}$$

where $\{p, q, r\}$ are consecutive pixels in the 3×1 and 1×3 neighborhoods described above. The function $\Gamma(\cdot)$ back projects the surface normal into the image plane. The first summation term defines the second derivative discretely applied to the three nodes. The second summation term creates a local surface normal by back projecting the three nodes and then retrieves the angle between the normal and the global up-vector. The difference from the resulting angle and the captured angle of polarization from the left camera is then sought to be minimized. As is well known, first-order priors tend to bias depth reconstructions to fronto-parallel surfaces. This is especially noticeable where images do not have strong texture all around the edge of the image. We tested both first-order and second-order smoothness priors and found that a second-order prior produces better results on our data sets, which is consistent with previous experience.

Note, that it is possible to also include the angle of polarization captured from the right camera into the smoothness term. However, we constrained ourselves to only include the angle of polarization from one viewpoint for the evaluation.

Graph Cut solution

The optimization is performed by a graph cut representation of the energy functional as quadratic pseudo-boolean optimization problem [27]. In this case, at each iteration a choice is made between a proposal disparity map and a current disparity map. Following Woodford et al [25], these proposals may be achieved by drawing randomly from a set of disparity values, or from different plane equations which are fitted to areas corresponding intensity segments in the input images. We found that constraining the unpolarized image regions to the disparity values that a SGBM [1] algorithm would produce works better than drawing all disparity values from a uniform distribution. Suitable proposals for the polarized regions can be acquired by enforcing planarity where the AOP-derived surface normal is constant. With the inclusion of visibility terms to the functional, the graph representation results to be non-submodular, so a global minimum solution can not be guaranteed [25] but the the global energy will decrease or remain constant with each iteration.

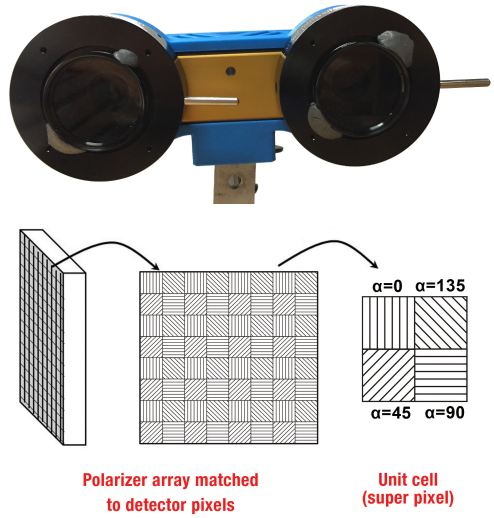

Fig. 4. The Bumblebee camera with adjustable linear polarization filters (up) and a depiction of the 4D polarcam micro-grid array (down, photo courtesy of 4D Technology).

VI. Results

A. Test setup

We used two image capture setups during our evaluation. Initially, we used a PointGrey Bumblebee BB2-08s2C stereo camera rig equipped with a fixture that holds linear polarizers in front of each lens (Fig. 4, up). Each linear polarizer is adjusted manually to suitable rotations, e.g. to $0°, 45°, 90°$, and $135°$. Thus, in this setup the scene elements have to remain static.

At a later stage, we acquired two IGV B0610M cameras that have been altered by 4D Technology Inc to bond a micro-grid array of linear polarizers, with filters rotated at $0°, 45°, 90°, 135°$ over each 2x2 block of pixels (Fig. 4, down). We have set the baseline spacing to 3.5 inches. The images are captured at a depth of 16 bits per pixel.

B. Sensor Noise Evaluation

To verify image quality before experiments, we evaluated the sensor noise of the IGV B0610M by capturing images of a diffuse wall and a linear polarization filter over a time span of 30 minutes. The exposure time and gain have been fixed to allow for a good contrast. We have computed the temporal standard deviation in both cases for each pixel and then averaged the values over the image. For the diffuse wall, we found a standard deviation of 0.7% for 16 bit images. For the imaging with the additional linear polarization filter, we found standard deviations of 0.9% (parallel to filter), 0.72% ($45°$ rotated from filter), and 0.3% (perpendicular to filter). The derived values for degree of polarization and for angle of polarization showed a standard deviation of 2% and $1°$, respectively. Further, we evaluated the sensor noise of the Pointgrey Bumblebee BB2-08s2C with the polarizer fixture attached, and found a standard deviation of 0.3%. These noise levels were satisfactory. We did not evaluate the noise levels of the Pointgrey camera for varying polarizer rotations

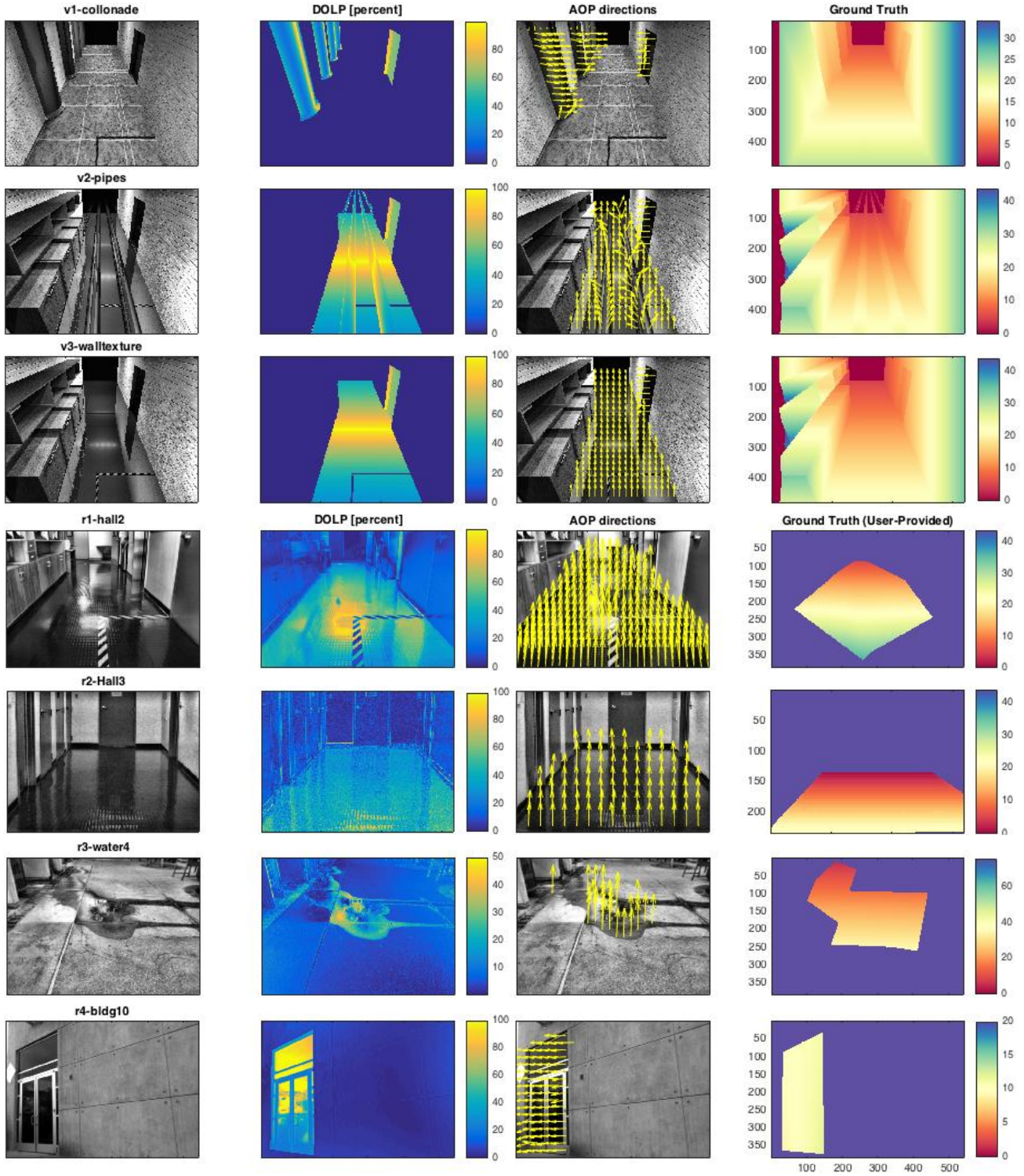

Fig. 5. Intermediate computation results and provided ground truth disparities on the virtual and real-world scenes. Left: the input image, polarizer at zero degrees, middle left : the Degree of Polarization in percent, middle right: the Angle of polarization visualized as vector plot, right: the ground truth disparity map. Note, that the AOP is only valid in image regions with high Degree of Polarization. The ground truth disparity maps for the real-world scenes are provided by user input.

of the external fixture, thus we did not compute the DOLP and AOP noise.

C. Scenes

We want to test our algorithms in scenes where shiny surfaces are present. In those scenes, conventional depth from stereo is assumed to be challenged. Thus we created three scenes in a modeling toolkit that showcase a large reflective portion of the scene in the stereo images. In scene v1-collonade, we design the surface material of the columns to be shiny. In scenes v2-pipes and v3-walltexture we design the floor to be shiny. The object geomtetry in both scenes is modeled after r1-hall2. Note, that in v2-pipes, the floor features two half-pipes whose computed AOP captures a large angular range. For the real world-scenes we have chosen reflective hallway floors, water puddles and windows. Most real world scenes show a subtle underlying surface texture in the shiny image regions (such as the PVC floor nobs). The scenes r1-hall2, r3-water4, and r4-bldg10 have been captured with the BB2-08s2C sensor, while the scene Hallway3 has been captured with the IGV B0610M sensor.

D. Comparison with state of the art

We compare the following there stereo matching algorithm to our approach on the datasets shown in Figure 5: SGBM [1]. Also, we run the original version of the graph cut algorithm, Woo2009 [25], on the data. We compute the absolute difference between the reconstructed disparity map I_{sm} and the ground truth disparity map I_{GT}. From the resulting difference map, we compute the mean $\mu_{|I_{sm} - I_{GT}|}$ within the pixel areas with nonzero degree of polarization, i.e. the reflective surface areas. We have chosen SGBM because it performs well for complex, diffuse scenery; though slower, it is still reasonably fast and real-time, low-power versions have been implemented with FPGAs and ASICs. We have conducted some experiments with the original, unaltered graph-cut algorithm as described in [25], and found that the photo consistency term governs the reconstruction in the specular areas. Therefore, the resulting depth maps show the typical artifact of virtual disparities from objects mirrored by the reflective surfaces. It is impractical to produce a pixel-accurate ground truth image I_{GT} for the real-world scenes with another sensor. However, we provide user-defined pseudo ground truth images $\hat{I}_{GT} = I_{user}$ by letting the user identify planar surfaces in both stereo images for each real-world scene. Each planar surface is attributed a normalized local coordinate system by solving a homography between the bounding box of each planar surface and a unit plane. The per pixel disparity $\hat{I}_{GT}$ is computed by subtracting the right from the left coordinate values for each planar surface individually. Unidentified areas are then excluded from the computation $|I_{sm} - I_{GT}|$. As preprocessing step we allow the adjustment of image intensity between the left and right image. Further, the AOP information in small pixel regions that are saturated for most of the polarizer rotations (e.g. at highlights) is replaced by interpolating with AOP information from neighboring pixels. The intermediate results for our algorithm, such as the DOLP and AOP are visualized in Figure 5. A side-by side comparison of the results from our approach and the other algorithms can be seen in Figure 6. The results of all algorithms applied to our test scenes are listed in Table I.

E. Discussion

1) Intermediate Results: Figure 5 shows the intermediate results for the virtual and real-world scenes. The second column visualizes the DOLP in percent. Note, that the surfaces, that have been rendered as purely diffuse return the same intensity value for any polarizer rotation. Hence, only the reflective areas have nonzero DOLP value. The close the incident angle to the Brewster angle, the higher the DOLP. The third column shows the AOP visualized as needle plot for the reflective areas only. The AOP follows the rotation of the projected surface normal in the image plane, see third column. The fourth column shows the ground truth disparity maps. Note, that the ground truth disparity value can be computed for each pixel from the geometry of the virtual scene. Note, that for the real-world scenes, the AOP only follows the rotation of the projected surface normal where the DOLP is significantly high. Since the geometry of the real-world scene elements is not known, the ground truth is provided by point correspondences, that have been clicked by a user.

2) Result Disparity Maps: The result disparity maps are shown in Figure 6. The second column shows the results using our algorithm. The third column shows the outcome using from Woo2009 [25], the fourth column shows the results for SGBM[1]. The last column reiterates the ground truth disparity maps for visual comparison. Further, in Table I, the mean absolute difference, the Out-2 and Out-4 percentage, denoting the percentage of pixels whose disparity value differs more than 2 or 4 units from the provided ground truth, are provided for all scenes (rows). For SGBM we also provide the density to indicate the percantage of pixels that have been assigned a valid disparity value. We applied a histogram adjustment before computing the Woo2009 result in some scenes (v1-collonades) where it improved the accuracy of the result disparity. It can be seen that the algorithms Woodford and SGBM that do not use polarization are challenged by reflections. For Scene V2, though, these algorithms seem to produce near-correct results in the pixel areas comprised by the pipes. However, the pipe-floor edge provides an important cue for assigning a close-to-correct disparity in that case. For the real-world scene r1-hall2, the hallway floor, which is a PVC material, is too specular to contribute to a meaningful depth reconstruction and the state of the art algorithms produce depth values in the pixel region comprised by the floor that are larger than they should be, corresponding to the virtual depth of the reflection. The same holds for the water puddle in r3-water4 and r4-bldg10. Our approach performs well on scenes V1, V2 and V3. It is challenged by the colonnades in V1, as surfaces normals varying along the camera's line of sight cannot be sufficiently distinguished by exclusively relying on the Angle

of Polarization. Note, that there is no back wall in the virtual hallway scenes and that the camera looks toward infinity at the and of the hallways. Our approach overcomes the effects introduced by the specular material in the real-world scene R1 and the floor is reconstructed convincingly. Further, our results for r3-water4 and r4-bldg10 are more accurate than the results from the algorithms that do not incorporate polarization information. It can be seen, that SGBM produces a stair-case looking result on r4-bldg10 which is reasoned by a number of slightly higher disparity gradients in the subpixel-accurate disparity map which invoke Mach-band effects.

VII. Conclusion

In this paper, we focused on 3-D perception of specular scenes, with eventual applications in mind to robot navigation in dynamic, indoor/outdoor environments. A wide variety of approaches have been taken in the past to this problem using passive and active sensors, including the use of structured illumination. Existing methods had a variety of limitations, in terms of maximum range, maximum surface slope, ability to operate with uncontrolled lighting, ability to estimate absolute range, or demonstrated suitability for mobile robot applications. Polarization phenomenology had not previously been integrated into a stereo vision algorithm. The advent of polarization cameras that acquire four channels of linearly polarized imagery simultaneously makes this prospect more practical than it used to be, and holds considerable potential to be useful for scenes with transparency as well as specularity. Accordingly, we developed a stereo algorithm that incorporates polarization-based surface normal constraints into typical stereo vision energy function; to improve performance on urban scenes with surfaces at a wide range of orientations, we also used a second-order smoothness prior. We leveraged existing software based on quadratric pseudo-boolean optimization to find depth maps that approximately minimize this cost function. Although the resulting implementation is still quite slow, results with synthetic and real polarization-stereo image sets show promise for this approach, and we believe that this sensor configuration warrants further research.

Acknowledgement

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(2) (2008) 328–341

[2] Blake, A.: Specular stereo. In: International Joint Conference on Artificial Intelligence. (1985)

[3] Blake, A., Brelstaff, G.: Specular stereo. In: International Conference on Computer Vision. (1988)

[4] Bhat, D.N., Nayar, S.K.: Stereo and specular reflection. International Journal of Computer Vision **26**(2) (1998) 91–106

[5] Li, Y., Lin, S., Lu, H., Kang, S.B., Shum, H.Y.: Multibaseline stereo in the presence of specular reflections. In: International Conference on Pattern Recognition. (2002)

[6] Jin, H., Yezzi, A.J., Soatto, S.: Variational multiframe stereo in the presence of specular reflections. In: International Symposium on 3D Data Processing, Visualization, and Transmission. (2002)

[7] Bonfort, T., Sturm, P.: Voxel carving for specular surfaces. In: International Conference on Computer Vision. (2003)

[8] Yang, R., Pollefeys, M., Welch, G.: Dealing with textureless regions and specular highlights: a progressive space carving scheme using a novel photo-consistency measure. In: International Conference on Computer Vision. (2003)

[9] Tsin, Y., Kang, S.B., Szeliski, R.: Stereo matching with linear superposition of layers. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2) (2006) 290–300

[10] Roth, S., Black, M.J.: Specular flow and the recovery of surface structure. In: Conference on Computer Vision and Pattern Recognition. (2006)

[11] Sizintsev, M.: Hierarchical stereo with thin structures and transparency. In: Canadian Conference on Computer and Robot Vision. (2008)

[12] Hebert, M., Krotkov, E.: 3D measurements from image laser radars: how good are they? International Journal of Image and Vision Computing **10**(3) (1992) 170–178

[13] Magda, S., Zickler, T., Kriegman, D.J., Belhumeur, P.N.: Beyond Lambert: reconstructing surfaces with arbitrary BRDFs. In: International Conference on Computer Vision. (2001)

[14] Mallick, S.P., Zickler, T., Kriegman, D.J., Belhumeur, P.N.: Beyond Lambert: reconstructing specular surfaces using color. In: Conference on Computer Vision and Pattern Recognition. (2005)

[15] Lamond, B., Peers, P., Ghosh, A., Debevec, P.: Image-based separation of diffuse and specular reflections using environmental structure illumination. In: International Conference on Computational Photography. (2009)

[16] de La Garanderie, G.P., Breckon, T.: Improved depth recovery in consumer depth cameras via disparity space fusion within cross-spectral stereo. In: Proc. British Machine Vision Conference. (2014) 417–1

[17] Wolff, L.B., Boult, T.E.: Constraining object features using a polarization reflectance model. IEEE Transactions on Pattern Analysis and Machine Intelligence **13**(7) (1991) 635–657

[18] Nayar, S.K., Fang, X.S., Boult, T.E.: Separation of reflection components using color and polarization. International Journal of Computer Vision **21**(3) (1997) 163–186

[19] Atkinson, G., Hancock, E.R.: Recovery of surface orientation from diffuse polarization. IEEE Transactions on Image Processing **15**(6) (2006) 1653–1664

[20] Herrera, S.E.M., Malti, A., Morel, O., Bartoli, A.: Shape-from-polarization in laparoscopy. In: International Symposium on Biomedical Imaging (ISBI), IEEE (2013) 1412–1415

[21] Photon-X. www.photon-x.com/

[22] Rahmann, S., Canterakis, N.: Reconstruction of specular surfaces using polarization imaging. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1., IEEE (2001) I–149

[23] Ngo Thanh, T., Nagahara, H., Taniguchi, R.i.: Shape and light directions from shading and polarization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 2310–2318

[24] Kadambi, A., Taamazyan, V., Shi, B., Raskar, R.: Polarized 3d: High-quality depth sensing with polarization cues. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 3370–3378

[25] Woodford, O., Torr, P., Reid, I., Fitzgibbon, A.: Global stereo reconstruction under second-order smoothness priors. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(12) (2009) 2115–2128

[26] Born, M., Wolf, E.: Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. CUP Archive (2000)

[27] Boros, E., Hammer, P., Sun, X.: Network flows and minimization of quadratic pseudo-boolean functions. In: Technical Report. RUTCOR (1991)

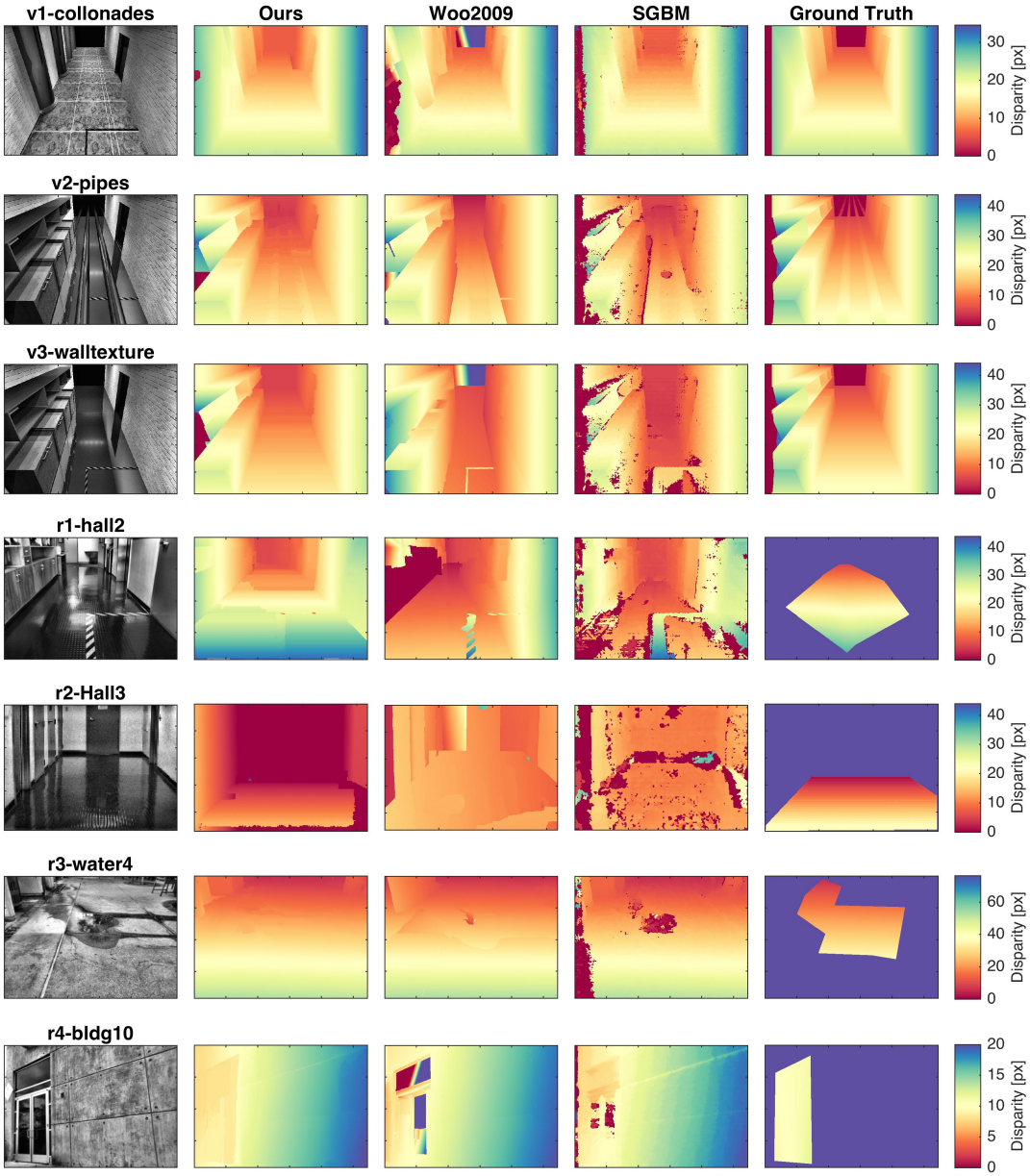

Fig. 6. The result disparity maps on both the virtual and the real-world scenes computed with our algorithms and with Woo2009 [25], SGBM [1].

	Ours			Woo2009[25]			SGBM[1]			
	Mean	Out-2 [Percent]	Out-4 [Percent]	Mean	Out-2 [Percent]	Out-4 [Percent]	Mean	Out-2 [Percent]	Out-4 [Percent]	Density [Percent]
v1-collonade	0.608	0.730	0.000	0.651	0.471	0.000	0.545	0.226	0.080	99.755
v2-pipes	0.426	0.424	0.086	1.319	6.206	3.392	1.165	4.913	2.107	98.743
v3-walltexture	0.528	0.041	0.000	4.566	21.012	15.312	5.936	20.938	17.249	98.590
r1-hall2	0.906	3.946	0.038	10.647	25.407	23.605	9.061	21.406	19.023	99.773
r2-hall3	3.713	12.065	5.213	37.217	35.368	35.368	7.834	30.202	25.127	89.604
r3-water4	0.302	0.043	0.000	3.004	4.560	4.287	2.355	3.235	2.823	99.408
r4-bldg10	1.749	7.344	0.000	8.790	7.425	5.433	2.795	13.151	2.369	99.445

TABLE I

THE RESULTS ON THE VIRTUAL SCENES RENDERED WITH OUR SIMULATOR AND THE REAL WORLD SCENES CAPTURED WITH OUR STEREO RIG. THE TABLE SHOWS THE MEAN $\mu_{|I_{sm} - I_{GT}|}$ OF THE ABSOLUTE DIFFERENCE BETWEEN THE GROUND TRUTH DISPARITIES AND THE COMPUTED DISPARITIES. FURTHER IT STATES THE AMOUNT OF DISPARITY PIXELS THAT ARE OFF MORE THAN 2 AND 4 UNITS FROM THE GROUND TRUTH. THE DENSITY OF VALID DISPARITY PIXELS IS LISTED FOR SGBM. SOME ALGORITHMS SHOW MINOR MEAN ABSOLUTE DIFFERENCES ([1]), BUT ALL ARE OUTPERFORMED BY OUR APPROACH.