Plane Surface Detection and Reconstruction using Induced Stereo Symmetry

Michel Antunes[1]
michel@isr.uc.pt
João P. Barreto[1]
jpbar@isr.uc.pt
Xenophon Zabulis[2]
zabulis@ics.forth.gr

[1] Institute of Systems and Robotics, Faculty of Sciences and Technology, University of Coimbra, 3030 Coimbra, Portugal

[2] Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Crete, Greece

Abstract

We propose an algorithm for the detection and reconstruction of plane surfaces using a new stereo approach dubbed *SymStereo*. SymStereo relies in symmetry analysis for recovering the 3D curve where a virtual cut plane intersects the scene structure. The result is a profile cut that resembles the one that would be obtained by a Laser Range Finder (LRF). The article shows that the framework is particularly well suited for piecewise-planar reconstruction using only a pair of calibrated views. Since the intersection of two planes is always a line, the 3D space is sampled by a discrete set of virtual planes and the line segments in the profile cuts are extracted. The plane surfaces are determined by grouping co-planar lines using a straightforward RANSAC procedure in the dual Plücker space. We test the algorithm in estimating the relative pose of the stereo rig with respect to planes with different textures. The results are highly accurate and, more importantly, the approach succeeds in situations where current stereo methods fail due to low and/or repetitive texture. We also report experiments in wide-baseline stereo images of complex scenes with multiple planes partially occluded by non-planar objects.

1 Introduction

Recovering 3D structure from stereo is a classical problem in computer vision with thousands of references in the literature. The approaches can be coarsely divided in two groups: *sparse stereo* (SS), that uses as input a sparse set of matching features across views; and *dense stereo* (DS), that performs dense data association between images by assigning to each pixel a disparity value. The former is used for recovering the camera parameters and/or obtaining a sparse point cloud model of the structure [12], while the latter usually assumes calibrated cameras and aims to achieve a complete 3D model of the scene [15]. The stereo approach presented in here, dubbed *SymStereo*, is somewhere between SS and DS. Given a calibrated stereo pair, we show how to reconstruct the curve where a virtual cut plane intersects the scene structure (see Fig. 1 (a) and (b)). While standard stereo relies on photo-consistency for matching pixels, we detect the image of the profile cut using signal symmetry analysis.

© 2011. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

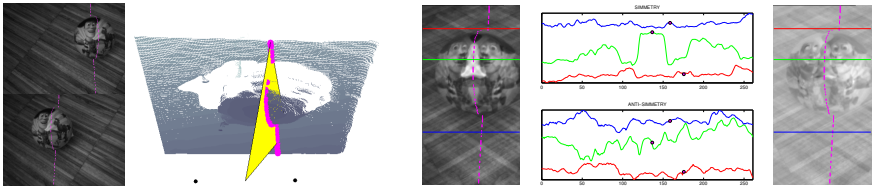
This is possible because, when the virtual plane is assumed to pass between the cameras, the homography mapping of one view into the other gives rise to a warped image that is mirrored with respect to the profile cut projection. Thus, the sum of the original and warped frames is symmetric with respect to this contour, and the image of the profile cut can be robustly determined by applying log-Gabor wavelets [10]. An important advantage of the symmetry cue is its global character that enables matching textureless image regions where photo-consistency measurements are ambiguous. Another interesting property of SymStereo is allowing to probe into the scene structure in a new and controlled manner. Given two images we are able to generate 3D data similar to the one that would be acquired by a LRF with the scan plane coincident with our arbitrary virtual cut plane. The article explores this unique feature for developing a new algorithm that, given two calibrated images, it detects and reconstructs planar surfaces in the scene.

Piecewise planar models have recently become popular for the reconstruction of man-made environments [6, 9, 14, 16, 19]. Using the strong planarity assumption as a prior for stereo overcomes difficulties caused by poorly textured surfaces and non-Lambertian reflections (e.g. windows). The achieved 3D models are perceptually pleasing and geometrically simple and, thus, their rendering, storage and transmission is computationally less complex. Unfortunately, locating planar surfaces in the scene for establishing plane hypotheses can be a very challenging task. As several approaches are described in the literature, the most recent and relevant are referenced. Bartoli obtains an initial sparse point reconstruction and applies a RANSAC-like algorithm for generating and scoring the plane hypotheses [2]. The approach lacks robustness to be used in complex scenes with several planar and non-planar surfaces. Assuming a Manhattan-world model, Furukawa et al. [6] reconstruct 3D patches from textured image regions, and uses their orientation to establish plane hypotheses. In [16], evidence about planar surfaces is collected using simultaneously point cloud reconstruction, estimation of vanishing lines, and sparse reconstruction of edges. Multiple cues and assumptions are also used in [14, 19] to find dominant surface orientations and perform plane-sweeping reconstruction in an optimal sweeping direction. Recently, Gallup et al. [9] proposed a modified RANSAC procedure for segmenting planar patches in dense multi-view stereo reconstructions. Although the above works address the problem in different manners, they all use multiple images to collect enough 3D evidence for supporting the plane hypotheses. Unlike these works, our article describes an algorithm for piecewise-planar reconstruction that uses only two calibrated images. The method can be useful for many applications, specially now that handheld stereo cameras, such as FujiFilm FinePix 3D and Sony Bloggie 3D, arrived to the consumer electronics market.

2 SymStereo: Stereo from Induced Symmetry

The plane sweeping algorithm was first introduced by Collins [4] and since then it has been widely used in dense depth estimation [20]. The basic idea consists in sampling the 3D space by a family of parallel virtual planes, back-project the images onto these planes, and find the locations where the back-projections are most similar. Ideally, these locations correspond to the intersection points of the plane with the imaged surfaces, which enables depth recovery. Recent forms of the algorithm avoid explicit back-projection and search for photo-consistent regions on the image plane after warping by the virtual plane homography [8].

Our approach for stereo reconstruction relates with plane-sweeping in the sense that it also samples the 3D space by virtual planes. However, there are two major differences: (i) we



(a) $\mathtt{I}$ and $\mathtt{I}'$ (b) 3D model (c) $\mathtt{I}_s{=}\mathtt{I}{+}\widehat{\mathtt{I}}$ (d) Epipolar Line (e) $\mathtt{I}_a{=}\mathtt{I}{-}\widehat{\mathtt{I}}$

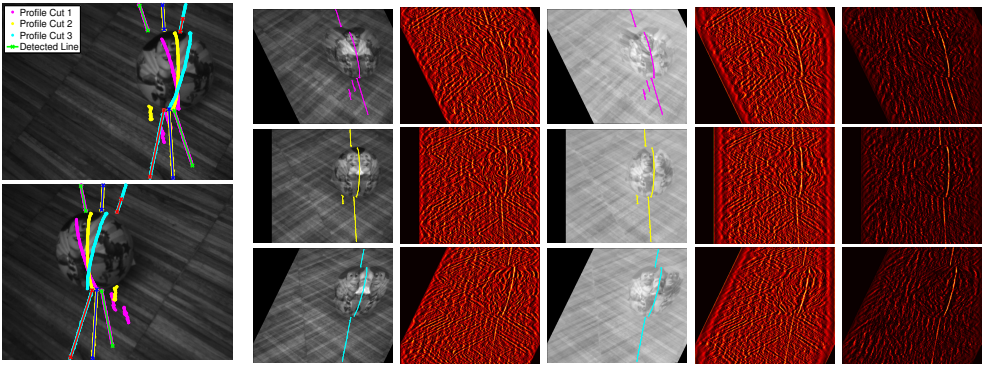
Figure 1: The left and right images are respectively $\mathtt{I}$ and $\mathtt{I}'$. The virtual cut plane in (b), passes between the cameras, and intersects the structure in a non-continuous 3D curve (the *profile cut*). $\mathtt{I}_s$ and $\mathtt{I}_a$ are symmetric and anti-symmetric with respect to image contour of the profile cut. They are rendered by adding and subtracting the original image $\mathtt{I}$ with $\widehat{\mathtt{I}}$, that is the result of warping $\mathtt{I}'$ by the plane homography. (d) shows the intensity level of $\mathtt{I}_s$ and $\mathtt{I}_a$ for three distinct epipolar lines (red, green and blue). The intersections with the profile cut projection can be identified with almost no ambiguity by searching the pixel location for which the top and bottom 1D-signals are respectively symmetric and anti-symmetric.

exclusively consider virtual planes that intersect the baseline in a point between the cameras; and (ii) instead of looking for photo-consistent regions after image warping, correspondences are established by detecting a mirroring contour using symmetry analysis. Works assuming multiple plane sweeping directions [8] consider virtual planes passing between the cameras a degenerate configuration to be avoided. The reason is as follows: Let $\mathtt{I}$ be the left stereo image and $\widehat{\mathtt{I}}$ be the result of warping the right image $\mathtt{I}'$ by the plane homography mapping. If the virtual plane crosses the baseline, then $\mathtt{I}$ and $\widehat{\mathtt{I}}$, instead of being photo-consistent, are reflected one with respect to the other around the pixel location where the plane cuts the structure. It is formally proved in [1] that this mirroring effect holds in general independently of the scene structure. The only singularity is the rare situation of *double nail ilusion* where the ordering constraint is violated [17].

Given the above, the sum of $\mathtt{I}$ and $\widehat{\mathtt{I}}$ yields an image signal $\mathtt{I}_s$ that is symmetric around the image contour of the profile cut (see Fig. 1(c)). In a similar manner, the subtraction of $\widehat{\mathtt{I}}$ from $\mathtt{I}$ gives rise to an image signal that is anti-symmetric at the exact same location (see Fig. 1(e)). This work explores these properties and proposes SymStereo, which reconstructs the profile cut by applying log-Gabor wavelet analysis (see [10] and [1]) for detecting the common pixel locations where $\mathtt{I}_s$ and $\mathtt{I}_a$ are respectively symmetric and anti-symmetric. As discussed in [1], the symmetry cue is affected by depth change, but the log-Gabor detection is robust to these deviations with SymStereo being able to accurately reconstruct extremely slanted surfaces.

3 Algorithm for Piecewise-Planar Reconstruction

This section describes an algorithm that uses induced symmetry from stereo for detecting and localizing planar surfaces in the scene. Since two planes always intersect into a line, we take advantage of this prior for achieving the goal by searching for 'line cuts'. A discrete set of N virtual cut planes intersecting the baseline in its midpoint is employed for probing into the 3D structure. The projection of the curve where each cut plane meets the scene is detected based on symmetry analysis. Line segments are extracted from the contour using a Hough transform, and the corresponding line cuts are reconstructed by back-projection into



(a) $\mathtt{I}$ and $\mathtt{I}'$ (b) $\mathtt{I}_s$ (c) $\mathtt{E}_s$ (d) $\mathtt{I}_a$ (e) $\mathtt{E}_a$ (f) $\mathtt{E}$

Figure 2: Image detection of profile cuts and segmentation into line cuts. Column (a) shows the input stereo pair. The rows in (b) to (f) refer to different cuts generated from three distinct virtual planes. The images of the reconstructed profile cuts are overlaid in (a) using different colors (magenta, yellow, and cyan). It can be observed that the repetitive texture of the floor gives rise to several local maxima in both $\mathtt{E}_s$ and $\mathtt{E}_a$, which precludes a correct detection of the relevant contour using a single type of energy. $\mathtt{E}$ is the pixel-wise multiplication of $\mathtt{E}_s$ and $\mathtt{E}_a$ and highlights pixel locations with both types of energy. We run ridge detection in $\mathtt{E}$ for locating the image of the profile cut (overlays at magenta, yellow, and cyan). The line segments in each contour detection are subsequently extracted using Hough transform (overlays at green, blue and red in (a)). The final line cuts are reconstructed by back projection on the virtual cut planes.

the virtual plane. Finally, all the recovered line cuts are clustered in a RANSAC stage using a co-planarity criterium. Each detected cluster gives rise to a plane hypothesis.

3.1 Processing each virtual cut plane

Let's assume a particular virtual cut plane Π, e.g. the yellow cut plane in Fig. 2. If Π intersects a planar surface, then the corresponding profile cut must contain a straight line segment (the blue line in Fig. 2).

Measuring Signal Symmetry - As described in Section 2, from each cut plane Π we render a pair of image signals $\mathtt{I}_s$ and $\mathtt{I}_a$, that are respectively symmetric and anti-symmetric around the contour where the profile cut is projected (see Fig. 1). The contour detection in these images requires to somehow quantify the signal symmetry and anti-symmetry along the epipolar lines. This is achieved by following the approach proposed by Kovesi in [10], that uses a Log-Gabor wavelet transform for generating a symmetry and anti-symmetry energy at every image pixel location (refer to [11] for a detailed description). These energies are denoted respectively by $\mathtt{E}_s$ and $\mathtt{E}_a$, and can be observed in Fig. 2(c) and (e).

Detecting the projection of the profile cut - It follows from the discussion of Section 2, that the contour that we aim detecting corresponds to an image pixel location for which it simultaneously occurs a maximum in $\mathtt{E}_s$ and $\mathtt{E}_a$. Thus, the joint energy $\mathtt{E}$ is generated by pixel-wise multiplication of $\mathtt{E}_s$ and $\mathtt{E}_a$. As observed in Fig. 2(f), this operation enables to discard many spurious local maxima that arise both in the symmetry and anti-symmetry energies. The projection of the profile cut is finally found by running standard ridge detection

in $\mathtt{E}$ [5]. The detection results are overlaid in the images of symmetry and anti-symmetry, $\mathtt{I}_s$ and $\mathtt{I}_a$, as well as on the stereo pair of Fig. 2 (each color refers to a different cut plane). There are some minor spurious detections that could be easily discarded with further refinement (e.g. to enforce the constraint of a unique detection per epipolar line).

Reconstructing line cuts - The ridge points detected in the joint energy $\mathtt{E}$ are used as input to a Hough transform that aims extracting line segments from the image of the profile cut. The result of this step is shown overlaid in the stereo pair of Fig. 2. The algorithm succeeds in finding the image of the line where each cut plane meets the floor.

3.2 Plane surface detection and reconstruction

At this stage we have a set of 3D lines computed from various cut planes. In order to cluster sets of co-planar lines, the contributions of all virtual planes are used as input in a RANSAC procedure. Each group of lines gives rise to a plane hypothesis contained in the scene.

Regarding the RANSAC implementation, there are aspects that we would like to point out: (i) the RANSAC search is carried in the dual 3D space with the lines parametrized in dual Plücker coordinates. Thus, the RANSAC seeks for sets of dual lines that are incident into a single point. This point is the dual representation of the plane hypothesis; (ii) Since a virtual cut plane cannot intersect the same plane hypothesis in two distinct lines, the sampling step always selects lines from different cut planes. (iii) After determining a plane model, the inlier set is calculated from the Euclidean distances of the plane to the lines.

4 Pose Estimation of a Single Plane Surface

This section assumes that the scene is dominated by a planar surface that might, or might not, be partially occluded by non-planar objects. We experiment in estimating the relative pose between the stereo rig and the plane for the case of different textures. The accuracy of the estimation is carefully evaluated with respect to ground truth data, and the performance of our algorithm is compared with two other strategies.

4.1 Data Acquisition and Estimation of Ground Truth

The first column of Tab. 1 shows different planar textures that were considered in the experiment[1]. These textures were divided into three categories: (i) *Texture:* corresponding to cases that are accessible for most stereo methods; (ii) *Middle texture/Repetitive pattern:* comprising examples that are challenging because of the existence of textureless regions, specularities, and/or repetitive patterns; and (iii) *Low texture:* corresponding to surfaces with very flat texture where most stereo methods are likely to fail. There is a fourth group concerning examples where the dominant plane is partially occluded by non-planar objects.

For each texture we acquired between 6 to 10 stereo pairs using a *Bumblebee* from Pt-Grey, with a baseline of 12cm and image resolution of 640×480. Each data set comprises images captured from different distances and relative rotation, with the depth and slant ranges being shown under the column of 'Dataset Info'. The *Bumblebee* was mounted on a tripod and the lighting conditions of the scenes were maintained unchanged. For each stereo pair used as input for the algorithms, we acquired an additional calibration pair with a checkerboard pattern attached to the planar surface. This calibration pair enabled to obtain accurate

[1] We only show results of 7 out of 12 textures. The complete experiment can be found in [4]

ground truth, by estimating the plane-to-image homographies from clicked points, and performing a suitable factorization of the result [12].

4.2 Compared Algorithms

We compare the following methods for recovering the relative pose:

SymStereo - This is our method that is described in the previous sections. The results shown in the table were obtained using $N = 21$ virtual cut planes, that intersect the baseline in the middle. From the 21 planes, 7 are vertical and equally spaced in azimuth, while the remaining 14 are obtained by slanting each vertical plane by an angle of $\pm 18.5^\circ$ [2].

Sparse Stereo (SS) - A set of sparse point correspondences between views is obtained using the SIFT algorithm [11]. Outliers are discarded based on the epipolar geometry derived from the camera calibration. The inliers are used as input in a RANSAC procedure that robustly estimates the pose of the dominant plane. Although we do not discuss herein the implementation details, it is important to emphasize that the RANSAC generates solutions from sets of 3 point correspondences. We use the fact that the rigid displacement between cameras is known, and estimate only the plane location. This is different from computing the homography and then factorize the result into rotation, translation, and plane location [12]. This assures a fair comparison between approaches, with all competing methods taking full advantage of the information that is available.

Dense Stereo (DS) - The surface is determined by fitting a plane to the 3D reconstruction obtained after dense stereo matching. The fitting is done in a robust manner using a straightforward RANSAC procedure. The stereo matching is performed using Normalized Cross-Correlation (NCC) with a 21×21 window. The size of the window was experimentally tuned in order to optimize accuracy and convergence of pose recovery in the evaluation images. Admittedly, such a large window is unlike to be the optimal choice for reconstructing scenes with depth discontinuities

It is worth mentioning that we experimented with a global MRF formulation [3] before deciding to use NCC. The reasons for this decision are twofold: (i) we did not notice significant differences in the results achieved with MRF and NCC. Typically, global approaches outperform local methods in situations of occlusion and discontinuity. Since such cases do not arise in our evaluation data, there was no clear advantage in choosing the MRF; and (ii) the MRF optimization is difficult to tune for converging both in close (20cm) and medium (1-2m) range images. Thus, we often observed situations of divergence in stereo pairs that the NCC would handle without a problem.

4.3 Results

The right side of Tab. 1 shows the average results for each data set using the different methods. The column 'Failure' indicates the percentage of stereo pairs for which the algorithm was unable to estimate a plausible relative pose. $A\%_{(B+C)}$ means that the in $B\%$ of the cases the RANSAC procedure did not converge, while in $C\%$ of the cases the estimation result was discarded because the error was above a pre-defined threshold (5% for translation, and 5° for rotation). The column 'RMS Error' concerns the accuracy of the relative pose estimation when the algorithm succeeds. For each data set we show the Root Mean Square (RMS)

[2] We experimented with a varying number of cut planes N and the results are available in [1]. For the case of 'Textured' surfaces, the algorithm provides accurate plane pose estimations even when the number of cuts is minimum ($N = 2$).

Table 1: Experiments in Relative Pose Estimation.

			Dataset Info	
Texture	Squares		Depth (mm)	$444-716$
			Slant (deg)	$2-40$
			# Images	10
Medium Texture	Wood2		Depth (mm)	$335-997$
			Slant (deg)	$39-61$
			# Images	8
	Sheet2		Depth (mm)	$321-800$
			Slant (deg)	$21-51$
			# Images	6
Low Texture	Wall1		Depth (mm)	226-530
			Slant (deg)	3-32
			# Images	6
	Wall2		Depth (mm)	$268-1105$
			Slant (deg)	$3-50$
			# Images	10
Object Occlusion	Ball1		Depth (mm)	444
			Slant (deg)	32
			# Images	3
	Ball2		Depth (mm)	490
			Slant (deg)	30
			# Images	3

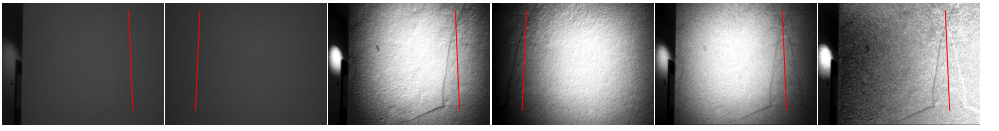
Method	Failure	RMS Error Trans.%	RMS Error Rot.(deg)	Average Ransac time(sec)
SymStereo	**0%**	**0.6**	**1.1**	**0.1**
DS	0%	0.4	1.0	0.63
SS	0%	0.2	0.8	0.05
SymStereo	**0%**	**0.8**	**1.8**	**0.13**
DS	38%(0+38)	2.3	2.1	12.68
SS	27%(13+14)	1.5	2.5	0.93
SymStereo	**0%**	**3.9**	**1.3**	**1.87**
DS	83%(0+83)	6.9	2.0	16.35
SS	83%(0+83)	1.8	4.8	3.57
SymStereo	**17%**(17+0)	**1.1**	**2.7**	**0.74**
DS	83%(0+83)	1.0	1.9	28.67
SS	100%(100+0)	-	-	0
SymStereo	**0%**	**0.9**	**3.9**	**0.76**
DS	80%(0+80)	0.6	3.1	28.84
SS	90%(90+0)	0.2	2.2	0.14
SymStereo	**0%**	**0.8**	**1.2**	**0.1**
DS	0%	0.5	0.6	1.62
SS	33%(0+33)	1	1.6	3.21
SymStereo	**0%**	**0.6**	**2.0**	**0.27**
DS	67%(0+67)	10.6	1.6	27.32
SS	100%(0+100)	-	-	32.75

value of the relative error in translation ('Trans'), as well as the angular error in orientation ('Rot'). Finally, the last column presents the RMS of the time interval that the RANSAC took to converge. These values were measured in a Matlab implementation, and are shown just for the sake of comparison between methods.

The three algorithms show a good behavior for the case of textured surfaces. The *SS* seems to be slightly more accurate, but the difference is negligible. More relevant is the fact that for the *DS* the RANSAC takes significantly longer to converge. This is not surprising because the cardinality of the input data is close to the image resolution. The time for clustering the co-planar points tends to grow exponentially with the number of outliers (e.g noisy 3D reconstruction, multiple surfaces, etc).

For the second texture category both *SS* and *DS* start to break. The failure of *SS* is explained by the problems in detecting enough keypoints that can be reliably matched. The *DS* has difficulties in handling repetitive patterns that are smaller than the correlation window ('Sheet2'), or have flat texture along the epipolar lines direction ('Wood2'). This causes ambiguity in the dense stereo matching that affects the 3D reconstruction and plane localization. The performance of the *SymStereo* is similar to the one observed for the 'Textured' category. For the 'Sheet2' there is a small decrease in translation accuracy, but this is mainly due to the lack of resolution for imaging a small pattern when the depth increases.

The third category comprises examples of almost textureless surfaces, where both *SS* and *DS* have notorious difficulties in delivering acceptable results. Surprisingly the *SymStereo* has almost no failures. The reason for this behavior is the global character of the symmetry analysis used in the detection of the mirroring contours. The virtual cut plane can intersect the surface in a completely flat region. However, and as shown in Fig. 3, nearby regions with some type of texture (e.g. small stain, scratch on the wall, etc) immediately contribute to define the correct symmetry when rendering the images $\mathtt{I}_s$ and $\mathtt{I}_a$. In this case, and as observed in the experiments, the accuracy of contour detection tends to decrease because the symmetry response is only at high wavelet scales. The experiments with objects on the top of the plane show that the *SymStereo* handles well partial occlusions.

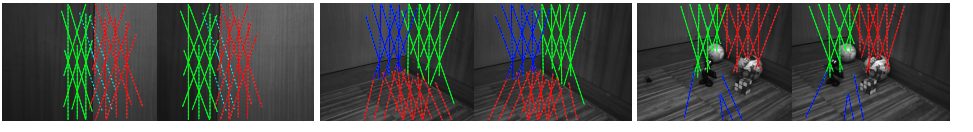


(a) $\mathtt{I}$ (b) $\mathtt{I}'$ (c) equalized $\mathtt{I}$ (d) equalized $\mathtt{I}'$ (e) equalized $\mathtt{I}_s$ (f) equalized $\mathtt{I}_a$

Figure 3: Line cut detection in a very low-textured surface (Wall1 in Tab.1). (a)-(b) shows the stereo images used as input to the algorithms. Fig. (c)-(d) and (e)-(f) are respectively the stereo pair and the symmetric and anti-symmetric signals after contrast enhancement (the equalization is performed only for visualization purposes). It can be observed that there is a thin scratch and very faded texture due to the wall rugosity. This is not enough for SS and DS to provide accurate reconstruction results. However, the scratch is sufficient to create a symmetry that enables SymStereo to correctly localize the line cut.

5 Detection and Reconstruction of Multiple Planes in Natural Scenes

This section shows results in detection and localization of multiple planar surfaces from a single stereo pair. The algorithm is similar to the one used in the previous experiment, with the difference that now we run RANSAC till convergence, remove the clustered co-planar lines, and repeat the procedure until generating a pre-defined number of plane hypothesis. In the absence of accurate ground truth, and since most images are scenes of man-made environments, the errors in plane orientation are estimated from likely relative parallelism and orthogonality.



(a) 2 Planes (b) 3 Planes (c) 3 Planes with Occlusion

Figure 4: Multiple plane detection in indoor scenes. Dashed lines correspond to the projection of the reconstructed line cuts, while solid indicates the line cuts that were assigned to one of the plane hypothesis by the RANSAC procedure (different colors signal different planes). The algorithm succeeded in the 3 situations, with an estimated relative orientation between planes of 92.1° for (a), 89.9° for (b), and 91.3° for (c) (average for the 3 planes).

Fig. 4 shows three stereo pairs of indoor scenes acquired with the *Bumblebee*. The *SymStereo* was run using $N = 21$ cut planes, and the localization of the plane surfaces was successful in all cases. The plane reconstruction results achieved with our stereo technique are suitable to be used as input data in "3D-plane-SLAM". This type of SLAM typically relies in LRF [13, 18] and, to the best of our knowledge, vision has never been used for this purpose.

Fig. 5 shows two outdoor scenes where the planes are partially occluded by non-planar surfaces. This is the type of environment targeted by the piecewise-planar reconstruction algorithms described in [2, 6, 9, 16]. While these methods require multiple views, our approach is able to reach competing results using only a stereo pair. For localizing the planar surfaces from the wide-baseline images of Fig. 5 we increased the number of cut planes from $N = 21$ to $N = 75$. After performing the detection in a manner similar to the previous exper-

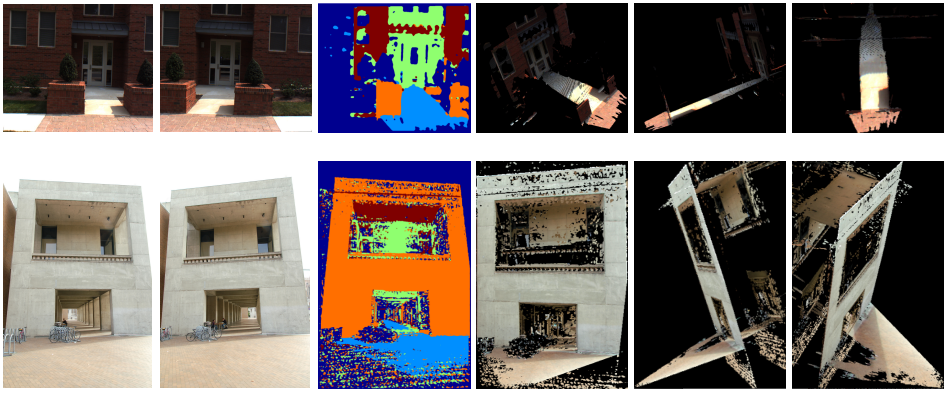


Figure 5: Wide baseline stereo of outdoor scenes. The first row concerns the *Out1* example [14], while the second row refers to the *Walls* example [7]. For each case we show the two stereo images, the left image with the pixels marked with different colors depending on the plane hypothesis to which they were assigned (similar to [9]), and three views of the obtained piecewise planar reconstruction. We detected and reconstructed 4 planes in each example. For *Out1* the average of the angle between the normals of the parallel vertical planes is 0.9°, and between the normals of the vertical planes and the floor is 88.1°. In *Walls*, the average angle between the normals of the vertical planes is 0.7°, and between the normals of the horizontal planes is 5° (the ground has a repetitive texture and the verandah top is relatively small). The mean angle between vertical and horizontal planes is 90.2°

iment, each pixel in the left image was assigned to the plane hypothesis that maximizes the photo-consistency with the right image. The result is shown in the middle column of Fig. 5. Pixels with low photo-consistency for all planes or that belong to the plane at infinity [9], were labeled as "discard" (dark blue). It is to expect that this labeling based exclusively in photo-consistency presents holes and poorly defined region borders. Such issue can be easily solved using MRF refinement [9, 16]. However, we chose not to do it in order to better assess the accuracy of our plane pose estimation.

The results of Fig. 5 show that it is possible to achieve accurate piecewise-planar reconstructions of everyday outdoor scenes using a single stereo pair. Remark that non-planar surfaces, like the bushes in *Out1* and the bikes in *Walls*, are correctly labeled despite of the depth proximity to detected planes. The fact that such segmentation is performed based only in photo-consistency is a qualitative evidence of accuracy.

6 Conclusions

We built on the the first stereo algorithm that exploits induced signal symmetry for determining the 3D curve where a virtual cut plane intersects the scene [10]. The approach was successfully applied in localizing planar surfaces using the fact that in this case the profile cuts are straight lines. The line cut assumption was used as a prior for improving the detection of the relevant mirroring contours. However, the experiments provide evidence that the symmetry cue alone is enough for reconstructing profile cuts of generic non-planar objects (see Fig. 2) . This new stereo framework can be advantageous in many computer and robot vision tasks other than planar surface reconstruction. First, while dense stereo techniques

are meant to estimate the disparity in the entire image, our method is designed to exclusively recover the depth along a predefined cut plane. This provides a new controlled manner for probing into the scene 3D structure, as illustrated by the approach used in here for detecting the planar surfaces. Second, as shown by the evaluation of Section 4, our approach succeeds in reconstructing surfaces with low and repetitive texture that can not be handled by standard stereo algorithms. The explanation is that we rely in symmetry cues that have a more global character than photo-consistency metrics used for the data association in the competing methods. Third, each reconstructed profile curve is similar to the cut that would be obtained by LRF with the laser plane aligned with our virtual cut plane. Thus, this new stereo technique can potentially avoid the use of a second sensor modality in systems that combine vision and LRF, or even replace the LRF in certain applications, such as 3D-plane-SLAM [13, 18], with benefits in terms of cost and flexibility.

Acknowledgment

Michel Antunes is grateful to the Portuguese Foundation for Science and Technology (FCT) by generous funding through the grant SFRH/BD/47488/2008. Joao P. Barreto would like to thank the Portuguese Agency for Innovation (ADI) that funded this research through the QREN co-promotion project n 2009/003512 sponsored by the Operational Program *Factores de Competitividade* and the European Fund for Regional Development.

References

[1] Michel Antunes and João P. Barreto. Plane surface detection and reconstruction using symstereo. Technical report, 2011. URL http://arthronav.isr.uc.pt/publications.php.

[2] A. Bartoli. A random sampling strategy for piecewise planar scene segmentation. *Computer Vision and Image Understanding*, 2007.

[3] Y.Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE PAMI*, 2004.

[4] Robert T. Collins. A space-sweep approach to true multi-image matching. In *CVPR*, 1996.

[5] D. Eberly. Ridges in image and data analysis. In *Series on Computational imaging and Vision*. Kluwer Academic Publishers, 1996.

[6] Y. Furukawa, B. Curless, S.M. Seitz, and R.S. Szeliski. Manhattan-world stereo. In *CVPR*, 2009.

[7] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE PAMI*, 2009.

[8] D Gallup, J.M. Frahm, P. Mordohai, Q.X. Yang, and M. Pollefeys. Real-time plane-sweeping stereo with multiple sweeping directions. In *CVPR*, 2007.

[9] David Gallup, Jean-Michael Frahm, and Marc Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. *CVPR*, 2010.

[10] Peter Kovesi. Symmetry and asymmetry from local phase. In *Australian Joint Conference on Artificial Intelligence*, 1997.

[11] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004.

[12] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[13] K Pathak, A Birk, N Vaskevicius, and J Poppinga. Fast registration based on noisy planes with unknown correspondences for 3-d mapping. *IEEE Transactions on Robotics and Automation*, 2010.

[14] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, and H. Towles. Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision*, 2008.

[15] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 2001.

[16] S. Sinha, D. Steedly, and R. Szeliski. Piecewise planar stereo for image-based rendering. In *ICCV*, 2009.

[17] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Symmetric stereo matching for occlusion handling. In *CVPR*, 2005.

[18] J Weingarten and R Siegwart. 3d slam using planar segments. *International Conference on Intelligent Robots and Systems*, 2006.

[19] Tomás Werner and Andrew Zisserman. New techniques for automated architectural reconstruction from photographs. In *ECCV*, pages 541–555, 2002.

[20] R. Yang and M. Pollefeys. Multi-resolution real-time stereo on commodity graphics hardware. In *CVPR*, 2003.