3-line RANSAC for Orthogonal Vanishing Point Detection

Jean-Charles Bazin and Marc Pollefeys
Computer Vision and Geometry Laboratory
Department of Computer Science
ETH Zurich, Switzerland
{jean-charles.bazin,marc.pollefeys}@inf.ethz.ch

***Abstract*— A wide range of robotic systems needs to estimate their rotation for diverse tasks like automatic control and stabilization, among many others. In regards of the limitations of traditional navigation equipments (like GPS and inertial sensors), this paper follows a vision approach based on the observation of vanishing points (VPs). Urban environments (outdoor as well as indoor) generally contain orthogonal VPs which constitutes an important constraint to fulfill in order to correctly acquire the structure of the scenes. In contrast to existing VP-based techniques, our method inherently enforces the orthogonality of the VPs by directly incorporating the orthogonality constraint into the model estimation step of the RANSAC procedure, which allows real-time applications. The model is estimated from only 3 lines, which corresponds to the theoretical minimal sampling for rotation estimation and constitutes our 3-line RANSAC. We also propose a 1-line RANSAC when the horizon plane is known. Our algorithm has been validated successfully on challenging real datasets.**

I. INTRODUCTION

This paper is dedicated to the estimation of the full rotation (3 degrees of freedom) for robotic applications. It is essential for a wide range of tasks like humanoid stabilization [1] and ground/aerial vehicle control [2] among many others. It also permits to greatly simplify the motion estimation and 3D reconstruction [3][4]. Traditional navigation equipments like GPS and IMU have well known limitations. For example, GPS cannot work in indoor environments or covered places (tunnel, underground) and IMU accumulates error over time.

With the advances of optics and the increase of computational power, a vision-based approach has gained in interest. A vision sensor is attached onto the robotic system and information is extracted from the images to estimate the camera rotation, and in turn, the rotation of the robotic system. Existing methods can be divided into three main categories. The first one relies on feature correspondence and epipolar geometry [5][6] or SLAM [7][8]. Whereas impressive results have been obtained recently, several important limitations and difficulties still exist: matching feature points is time consuming and a complicated task in real world applications and some epipolar models are scene-dependent (e.g. essential matrix is unstable in planar environment). Moreover they are overkill in the sense they estimate the complete motion (rotation plus translation), although only rotation might be needed (e.g. airplane orientation stabilization).

The second category refers to the feature correspondence-free methods. For example, [9] proposed an appealing probabilistic approach but assumes that the number of features is known and does not treat outliers and occlusion. [10] works in the frequency domain by spherical Fourrier transform, and thus is computationally expensive. Moreover, it is sensitive to translation and dynamic environment.

The third category takes into consideration the points at infinity since their motion is translation invariant. [11] developed an appearance-based approach and showed interesting results. However it suffers from several limitations: it estimates only one angle, is sensitive to occlusion, and assumes that the tracked image part is far away, the car follows a planar motion and the camera is perfectly vertical. An important set of points at infinity is the vanishing points (VP), that are the intersection of the projection of world parallel lines in the image. The approach presented in this paper belongs to this third category and is based on VPs.

The remainder of this paper is divided into three main parts. In section II, we review the existing works on rotation estimation based on lines and VPs. Then we introduce our new approach, namely the 1- and 3-line RANSAC, in section III. Finally, we present some experimental results.

II. EXISTING WORKS

Urban environments generally exhibit numerous lines that are either parallel or orthogonal to the gravity direction, which leads to orthogonal VPs (the so-called Manhattan world). Maintaining this orthogonality constraint is of key importance to respect the scene structure [12][13][14]. Given a set of lines extracted in images, many methods have been proposed for clustering these lines into sets of world parallel lines and computing their associated VPs, which, in turn, permits to compute the rotation. The common goal is to search for the VPs supported by the highest number of lines. Existing methods can be divided into four categories.

The first category is based on the Hough transform (HT) [15][16][17][18]. The direction of the intersection of each line pair is computed and accumulated in the angular bins. The bin containing the highest number of entries corresponds to the dominant VP. This approach is sensitive to the quantization level of the bins and might lead to multiple detections [19]. Moreover, since the VPs are detected independently, it does not directly impose the VP orthogonality constraint.

The second category alternates between the line clustering and VP estimation steps [20][12], generally following a expectation-maximization (EM) algorithm. Given an initial VP, EM alternates between performing an expectation (E) step, which computes the expectation of the line clustering evaluated for the current VP, and a maximization (M) step,

which computes the VPs given the data clustering found at the E step. This process iterates with the new VPs until convergence. These alternating methods heavily rely on the initialization and do not fulfill our real-time requirement.

The third category relies on a (quasi-)exhaustive search. For example, [21] samples the rotation search space to determine the rotation maximizing the number of clustered lines. This method can obtain very satisfying results and is able to enforce the VP orthogonality. However it depends on the sampling rate and has to process a lot of samples for fine sampling or large search space.

The fourth category refers to the popular RANSAC framework [22]. It is a simple but efficient general method to distinguish inliers and outliers and also estimate the underlying dominant model. [23][24][25], among many others, applied it for VP estimation. Two lines are randomly selected to create a VP hypothesis and the number of lines passing by this hypothesized VP is counted. This procedure is repeated during several iterations and outputs the VP (and the associated line clustering) that maximizes the number of line inliers. To detect multiple VPs, RANSAC can be sequentially applied on the remaining outliers [25] but cannot permit to impose the VP orthogonality. We refer to this technique as "independent RANSAC". J-linkage algorithm [26] applied by [27] obtained interesting results but is too much computationally expensive for real-time robotic applications. Finally, [28] presented a polynomial approach to impose the VP orthogonality but can only consider algebraic cost and does not run in real-time.

III. PROPOSED APPROACH

We now propose our new approach. Our two main objectives are VP orthogonality and real-time. First, concerning the orthogonality, let notice that if a method does not return orthogonal VPs (e.g. HT or independent RANSAC), an orthonormalization step based on the Frobenius norm to find the "nearest" orthogonal solution [29] is generally applied. Since this is performed after the model estimation, it decreases the quality of the solution. A similar phenomena can be observed by the 7-point algorithm for the fundamental matrix computation in which the singular values are corrected afterwards [30]. Therefore it is important to explicitly enforce the VP orthogonality constraint directly when the model is estimated.

Second, given the large proportion of line outliers, a robust estimator is necessary. For real-time applications, RANSAC [22] is a popular and appropriate choice. In the following, we explain our approach to incorporate the orthogonality constraint directly into the model estimation step of RANSAC. We start by reminding the spherical representation, then we present the line parametrization adopted in this paper and finally introduce our 3-line and 1-line approaches.

A. Spherical representation

Gaussian sphere is a convenient way to represent an image when the intrinsic calibration parameters of the camera are known. It has been used since [15] and has been generalized to the concept of equivalent sphere to represent images acquired by various types of cameras such as pinhole, fisheye, omnidirectional and catadioptric [31]. Interested readers are invited to refer to [31][32] for details about the projection equations between the original image and its spherical representation.

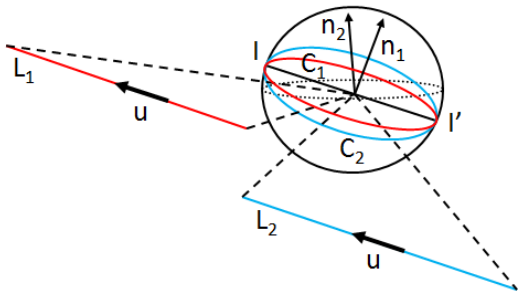


Fig. 1. Equivalent sphere projection: a world line (L_1) is projected onto the sphere as a great circle (C_1) and the projection of parallel lines (L_1 and L_2) intersect in 2 antipodal points (I and I').

A world line L_i is projected onto the equivalent sphere as a great circle which is represented by a unit normal vector n_i. The great circles of world parallel lines intersect in two antipodal points, as illustrated in Fig 1. They correspond to the vanishing point and are computed by $v = n_i \times n_j$ where n_i and n_j are the normal vectors of the great circles corresponding to two world parallel lines.

B. Line extraction

We applied the generalized polygonal approximation of [4] to extract the lines. For completeness and self-readability, we briefly explain it in the following. The method starts by extracting edges and building chains of connected edge pixels. Then these chains are projected onto the equivalent sphere and the great circle constraint (cf section III-A) is tested. For each chain, this test is performed by computing the geodesic (angular) distance between the chain points and the great circle whose normal vector is obtained by $n = P_1 \times P_M$ (where P_i is the i^{th} point of the chain in the sphere and M is the chain length). If the maximum distance is lower than a threshold, then the chain is considered a line. Otherwise, we split the chain at the furthest point and apply the same procedure for the two sub-chains until they are considered a line or their length is too short. We finally obtain a list of normals corresponding to the great circles of the detected lines.

C. 3-line RANSAC

RANSAC [22] processes as follows: random selection of data, model estimation and count of the inliers verifying this model. This process is repeated a certain number of iterations and outputs the model leading to the highest number of inliers. In the following, we explain our approach to enforce the VP orthogonality directly into the model estimation step.

A first important issue is the number of required samples that are randomly selected to estimate the model since the number of RANSAC iterations (and thus the algorithm complexity) exponentially grows with the number of required samples [30]. This refers to the notion of "minimal sampling"

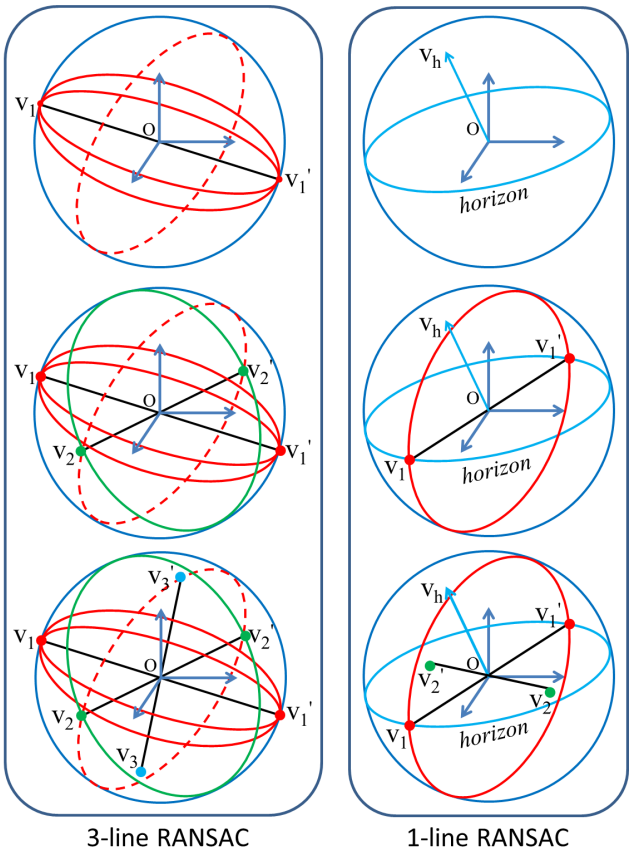


Fig. 2. Illustration of our proposed 3-line (left) and 1-line (right) RANSAC algorithms. cf text for detailed explanations. Best seen in color.

(also called "minimal solution"). A considerable amount of energy has been devoted to this topic in the computer vision community, such as [33][34][35][36], generally leading to mathematically appealing but computationally expensive polynomial systems. In contrast, we will show that our approach is both simple and computationally cheap.

Complete rotation has three DOF and each line provides one constraint. Therefore the minimal solution for rotation estimation is three lines. Our approach uses this minimal sampling and its process is illustrated in Figure 2-left. Each RANSAC iteration starts by randomly selecting three lines. The first two lines (shown in red) intersect at a VP v_1 (and its antipodal point v_1'). The great circle of normal v_1 (shown in dashed red) corresponds to the definition set of all the possible orthogonal VPs. The third line (shown in green) intersects this set at v_2 (and its antipodal point v_2'), which defines the second VP. Finally, the third VP v_3 is computed by the cross-product of v_1 and v_2. This last step is not necessary if only two VPs have to be extracted. Like for the traditional point-based epipolar methods [30], the process can be further accelerated and more efficient by privileging some geometric configurations of samples: lines with different orientations and a third line distant from v_1. Finally, it is interesting to note that the main computation is performed by simple cross and dot products, which are computationally cheap operators.

To emphasize that the orthogonality constraint is directly

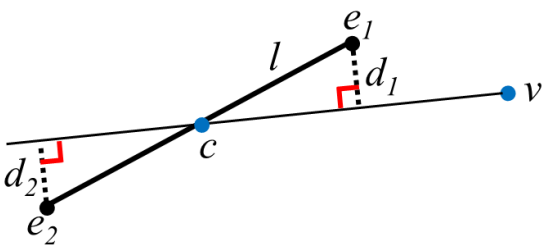


Fig. 3. Distance between a line l (whose middle point is c and endpoints are e_1 and e_2) and a vanishing point v in the image space.

incorporated in the model estimation step of RANSAC and in analogy to the naming of the traditional epipolar-based methods (e.g. the 4-point algorithm for homography [30], the 5-point algorithm for the essential matrix [36] and the 7- and 8-point algorithm for the fundamental matrix [30]), we refer to our method as "3-line RANSAC".

D. 1-line RANSAC

Inspired by the 1-point RANSAC of [37] in the context of car trajectory estimation, we developed a 1-line RANSAC to obtain the orthogonal VPs and the rotation. We alleviate the assumption of a planar non-holonomic motion of [37] by the knowledge of the horizon plane. This plane can be known apriori (e.g. camera vertically aligned), manually estimated or automatically computed by any existing techniques like sky/ground segmentation [38][39].

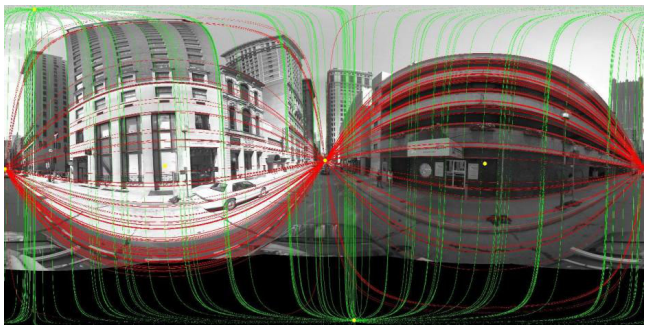
Figure 2-right illustrates our approach. The vector v_h represents the normal of the horizon plane (shown in blue), which generally corresponds to the vertical/gravity direction. Each RANSAC iteration starts by randomly selecting one line. This line (shown in red) intersects the horizon at a VP v_1 (and its antipodal point noted v_1'). Finally, the third VP v_2 is computed by the cross-product of v_1 and v_h.

E. Inlier definition

Given the VPs obtained by the 1- and 3-line methods, the next RANSAC step is to count the number of inliers. A line l is considered an inlier when its distance to the VP v is lower than a residual threshold [22]. This distance can be computed in several ways. One way is to compute the geodesic distance on the sphere: i.e. $|\arcsin(n \cdot v)|$ where n represents the great circle normal of the line l. Another way is to compute the distance in the image space: for example the average orthogonal distance from the endpoints of the line l to an auxiliary line $\tilde{l}$ defined as the line passing through v and minimizing this average orthogonal distance (or passing through v and the middle point of the line l), as illustrated in Fig 3 [30]. It is also possible to consider the maximal orthogonal distance [27].

IV. EXPERIMENTAL RESULTS

This section presents experimental results obtained by the proposed approach. Our algorithm accepts as input a set of lines (cf section III-B) and returns, in a fully automatic way, the detected VPs, the line clustering (i.e. which line belongs to which VP) and the 3D rotation. Additional results are



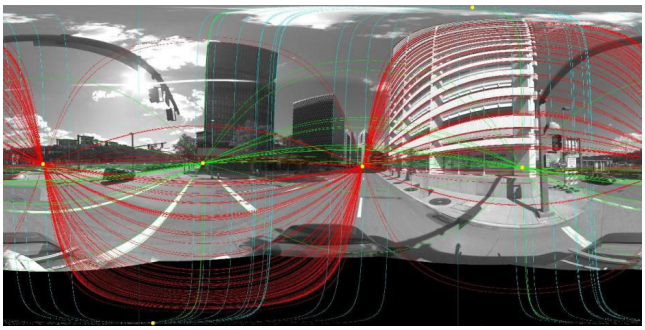


Fig. 4. Representative examples of line clustering and vanishing point extraction by the proposed 3-line RANSAC on Google Street View images. Each conic corresponds to a detected line and all parallel lines have the same color. For a better visualization, the conics have been enlarged and the supporting line pixels are not displayed.

available on the authors' website. For all the experiments shown in this paper, the number of RANSAC iterations for our approach is automatically computed with a safe outlier ratio of 70%, a guaranteed accuracy of 99% and a minimal sampling of 1 and 3, respectively for the 1- and 3-line RANSAC, which leads to only 13 and 169 iterations [30]. In addition, the involved operations are computationally cheap, and thus the proposed approach can run in real-time.

A. Google Street View

We applied our algorithm on real omnidirectional images issued from the Google Street View dataset[1]. The sequences have been acquired by an omnidirectional camera attached on the roof of a car driving in the city of Pittsburgh, USA. The projection of the input image onto the equivalent sphere is performed by linear mapping like in [40]. Because of the distortions inherent to the wide field of view, the world lines appear as not straight lines, but as conics in the image. Figure 4 illustrates some representative results of line clustering and VP detection obtained by the proposed 3-line RANSAC.

Figure 5 shows results obtained by the proposed 1-line RANSAC. Since the camera of the Google Street View dataset is aligned with the vertical, the normal of the horizon is simply $(0, 0, 1)$. The corresponding horizon line is displayed in blue. Note that this calibration step needs to be done only once. Experiments showed that the proposed 1-line RANSAC provided satisfying results and is perfectly suitable for real-time applications.

[1]copyrighted and kindly provided by Google.

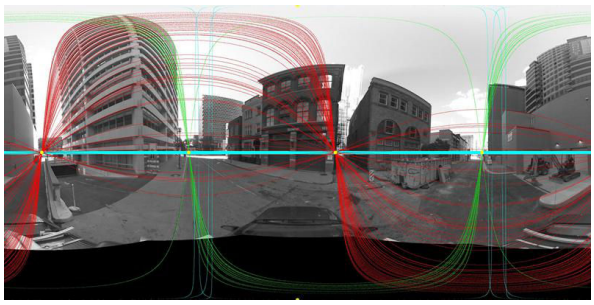


Fig. 5. Line clustering and vanishing point extraction by the proposed 1-line RANSAC. Same color code as in Figure 4. The blue line represents the horizon line.

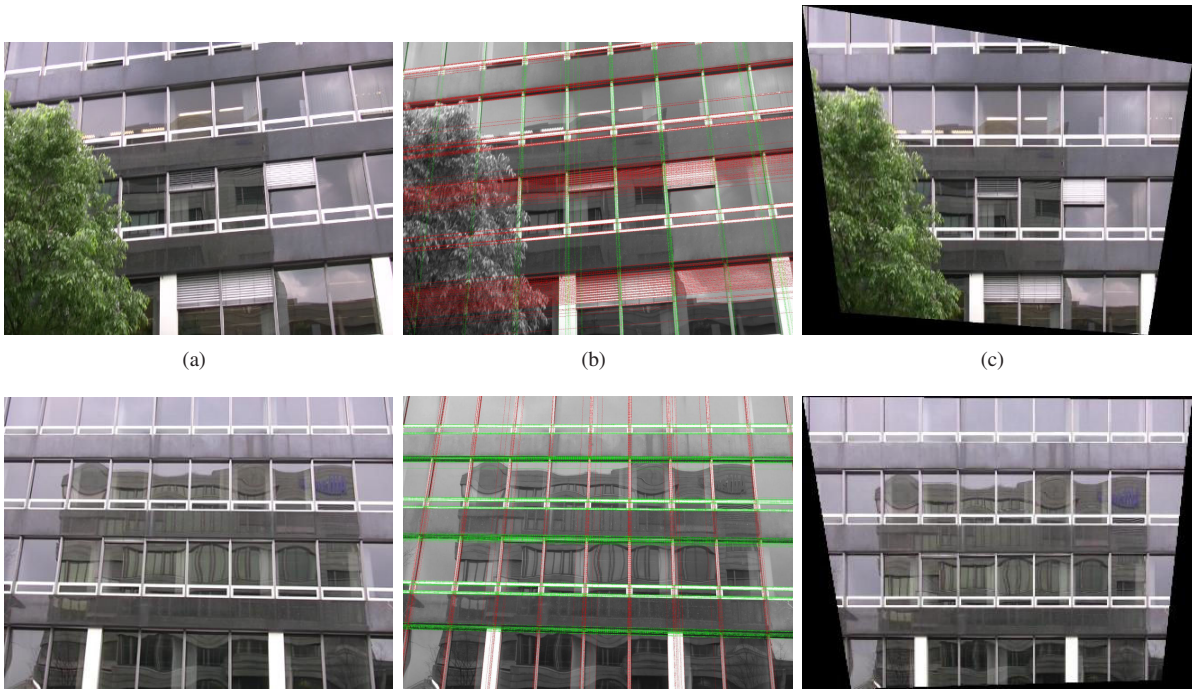
B. Reflective facade

We also applied our approach to two challenging sequences of a building facade composed of multiple reflective windows acquired by a hand held camera. Because of the large portion of reflections in the image, feature tracking (e.g. Harris corners+KLT [41][42] or SIFT [43]) cannot work reliably and thus visual odometry methods [5][6] are likely to fail. Figure 6-(b,e) presents some representative results. It shows our method works also when only two VPs are visible, without any difficulties. Since the rotation ground truth is not available, we performed fronto-parallel image rectification for verification (cf Fig 6-(c,f)). We obtained visually appealing results for the whole sequences: the rectified images are stable and the horizontal and vertical lines of the facade are now perfectly aligned with the image axis, which demonstrates the robustness and accuracy of our approach.

C. Comparison

We compared the algorithm complexity of the sampling approach (noted SA) of [21] and the proposed 1- and 3-line RANSAC. SA was selected because it can directly enforce the VP orthogonality. For SA, we considered the sampling rates T of $0.1°$ (fine) and $0.5°$ (coarse) with a search space size varying between $0.5°$ (very small) and $10°$ (intermediate) for each angle. As explained above, the number of RANSAC iterations was automatically computed to 13 and 169, respectively for the proposed 1- and 3-line approaches. The complexity comparison is shown in Figure 7. Concerning SA, the number of iterations greatly increases with the search space size, especially when the sampling rate is finer. As expected, the complexity of our RANSAC algorithms is always smaller than for SA, except when a coarse sampling ($0.5°$) and a small search space (less than $1.5°$) are considered, which is not recommended for real examples: the coarse sampling likely misses the solution, which leads to non-accurate results, and the small search space assumes a very limited inter-frame motion.

In terms of line clustering, Figure 8 shows a representative comparison of the results obtained by "independent RANSAC" (cf section II), the sampling approach of [21] (with a fine sampling rate of $0.1°$) and the proposed 3-line RANSAC. It confirms the fact that independent RANSAC can capture bundles of lines, but not especially orthogonal bundles. The proposed method and the sampling approach



(a) (b) (c)

(d) (e) (f)

Fig. 6. Representative results obtained by our 3-line RANSAC for two sequences of reflective facade (top and bottom rows). (a,d): original image; (b,e): line clustering and VP detection, (c,f): fronto-parallel image rectification.

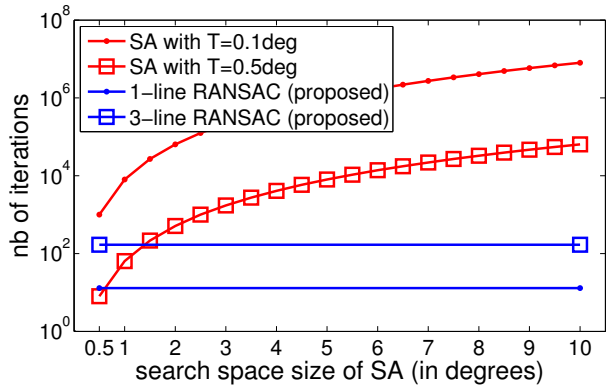


Fig. 7. Comparison of the algorithm complexity of the sampling approach (noted SA) [21] with different parameters and the proposed 1- and 3-line RANSAC. The x-axis represents the search space size for SA and the y-axis corresponds to the number of iterations in logarithmic scale.

return similar clustering results, the difference being that our algorithm requires much less iterations (cf Fig 7). For additional comparison between the 3-line RANSAC and SA, we measured the efficiency of the algorithms in terms of number of clustered lines. Results are presented in Figure 9. As expected, the SA efficiency is improved with a finer sampling rate. Moreover it shows our 3-point RANSAC outperforms SA. The low quality of the results of SA around the frame 120 is due to the fact that the car had a sudden rotation, larger than the search space size of 10°. In contrast, our method does not assume any motion smoothness and can handle this situation without any difficulties.

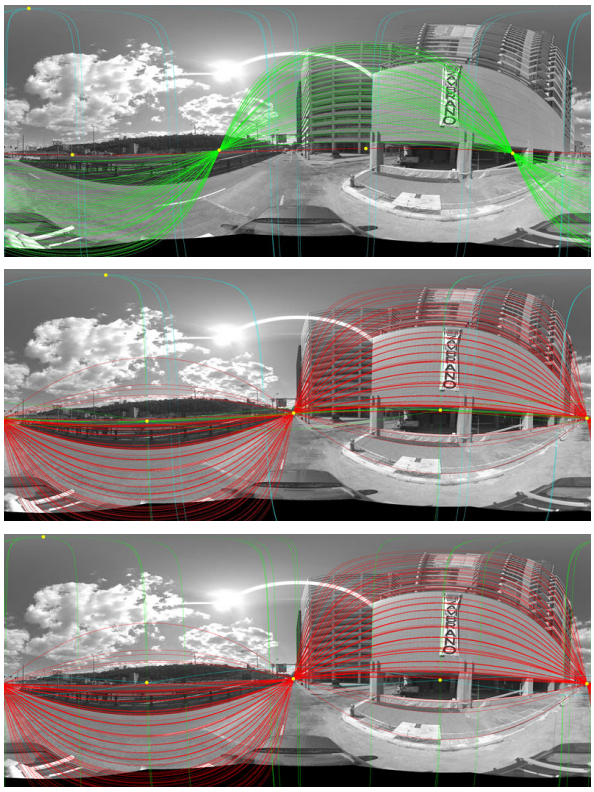


Fig. 8. Comparison of line clustering and VP detection by independent RANSAC (top), the sampling method of [21] (middle) and the proposed 3-line RANSAC (bottom).

V. CONCLUSION

In this paper, we have introduced a simple-but-powerful fully automatic 3-line RANSAC algorithm. The VP or-

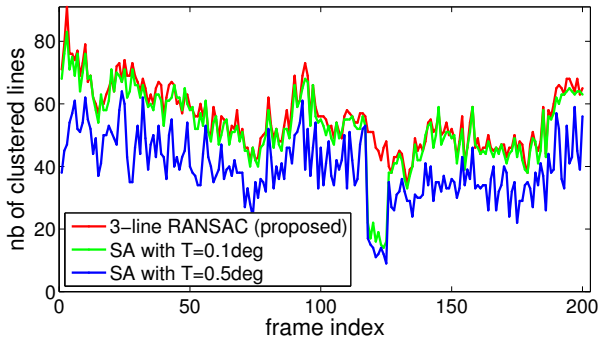


Fig. 9. Comparison of the number of clustered lines for 200 consecutive frames of a Google Street View sequence by the proposed 3-line RANSAC and the sampling approach (noted SA) [21] with the sampling rates $0.1°$ and $0.5°$ with a search space size of $\pm 10°$ for each angle.

thogonality constraint is directly enforced during the model estimation step of the RANSAC procedure and uses just 3 lines as the minimal sampling. It provides several important advantages: simultaneous estimation of multiple VPs, orthogonality constraint and low complexity for real-time processing, which constitutes the main contribution of our work. Our second contribution is a related 1-line RANSAC for situations where the horizon plane is known. Our approach has been successfully applied on challenging real data such as the Google Street View dataset and reflective facades.

ACKNOWLEDGMENT

This research has been funded in part by the ECs Seventh Framework Programme (FP7/2007-2013) / ERC grant #210806 4D Video. This research, which has been partially carried out at BeingThere Centre, is also supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

REFERENCES

[1] M. Finotto and E. Menegatti. Humanoid gait stabilization based on omnidirectional visual gyroscope. In *Workshop on Humanoid Soccer Robots (Humanoids'09)*.
[2] E. Rondon, L.-R. Garcia-Carrillo, and I. Fantoni. Vision-based altitude, position and speed regulation of a quadrotor rotorcraft. In *IROS'10*.
[3] M. Brand, M. Antone, and S. Teller. Spectral solution of large-scale extrinsic camera calibration as a graph embedding problem. *ECCV'04*.
[4] J.-C. Bazin, C. Demonceaux, P. Vasseur, and I.S. Kweon. Motion estimation by decoupling rotation and translation in catadioptric vision. In *CVIU'09*.
[5] A. Akbarzadeh and et al. Towards urban 3D reconstruction from video. *3DPVT'06*.
[6] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. *ICCV'09*.
[7] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *ICCV'03*.
[8] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich. The vSLAM algorithm for robust localization and mapping. In *ICRA'05*.
[9] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun. Structure from motion without correspondence. In *CVPR'00*.
[10] A. Makadia and K. Daniilidis. Rotation recovery from spherical images without correspondences. *PAMI'06*.
[11] D. Scaramuzza and R. Siegwart. Appearance guided monocular omnidirectional visual odometry for outdoor ground vehicles. In *IEEE Transactions on Robotics*, 2008.
[12] P. Denis, J. H. Elder, and F. J. Estrada. Efficient edge-based methods for estimating manhattan frames in urban imagery. In *ECCV'08*.
[13] R. Cipolla, T. Drummond, and D. Robertson. Camera calibration from vanishing points in images of architectural scenes. In *BMVC'99*.
[14] J. Kosecka and W. Zhang. Video compass. In *ECCV'02*.
[15] S. T. Barnard. Interpreting perspective image. *Artificial Intelligence Journal*, 1983.
[16] L. Quan and R. Mohr. Determining perspective structures using hierarchical Hough transform. *Pattern Recognition Letters*, 1989.
[17] M. J. Magee and J. K. Aggarwal. Determining vanishing points from perspective images. *Proc Computer Vision, Graphics and Image Processing*, 1984.
[18] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons. The cascaded Hough transform as an aid in aerial image interpretation. In *ICCV'98*.
[19] J. A. Shufelt. Performance evaluation and analysis of vanishing point detection techniques. *PAMI'99*.
[20] M. E. Antone and S. J. Teller. Automatic recovery of relative camera rotations for urban scenes. In *CVPR'00*.
[21] J.-C. Bazin, C. Demonceaux, P. Vasseur, and I.S. Kweon. Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. *IJRR'12*.
[22] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of the ACM*, 1981.
[23] D. G. Aguilera, J. G. Lahoz, and J. F. Codes. A new method for vanishing point detection in 3D reconstruction from a single view. In *Proceedings of ISPRS comission*, 2005.
[24] H. Wildenauer and M. Vincze. Vanishing point detection in complex man-made worlds. In *International Conference on Image Analysis and Processing (ICIAP'07)*.
[25] S. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys. Interactive 3D architectural modeling from unordered photo collections. In *SIGGRAPH Asia'08*.
[26] R. Toldo and A. Fusiello. Robust multiple structures estimation with J-linkage. In *ECCV'08*.
[27] J. P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV'09*.
[28] F. M. Mirzaei and S. I. Roumeliotis. Optimal estimation of vanishing points in a Manhattan world. In *ICCV'11*.
[29] G. Golub and C. Van Loan. *Matrix Computation*. Johns Hopkins University Press, Maryland, third edition, 1996.
[30] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.
[31] X.H. Ying and Z.Y. Hu. Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model? In *ECCV'04*.
[32] C. Mei. Laser-augmented omnidirectional vision for 3D localisation and mapping. In *PhD Thesis*, 2007.
[33] M. Byrod, Z. Kukelova, K. Josephson, T. Pajdla, and K. Astrom. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. *CVPR'08*.
[34] D. Nistér and H. Stewénius. A minimal solution to the generalized 3-point pose problem. *Journal of Mathematical Imaging and Vision*, 2006.
[35] H. Stewénius, C. Engels, and D. Nistér. An efficient minimal solution for infinitesimal camera motion. *CVPR'07*.
[36] D. Nistér. An efficient solution to the five-point relative pose problem. In *PAMI'04*.
[37] D. Scaramuzza. 1-point-RANSAC structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *IJCV'11*.
[38] S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M.Waszak. Vision-guided flight stability and control for micro air vehicles. *Advanced Robotics*, 2003.
[39] C. Demonceaux, P. Vasseur, and C. Pégard. Robust attitude estimation with catadioptric vision. In *IROS'06*.
[40] A. Banno and K. Ikeuchi. Omnidirectional texturing based on robust 3D registration through Euclidean reconstruction from two spherical images. *CVIU'09*.
[41] C. Harris and M. Stephens. A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference*, 1988.
[42] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI'81)*.
[43] D. Lowe. Distinctive image features from scale-invariant keypoints. In *IJCV'03*.