An adaptive descriptor for uncalibrated omnidirectional images - towards scene reconstruction by trifocal tensor

Ming Liu, Bekir Tufan Alper, Roland Siegwart
Autonomous Systems Lab, ETH Zurich, Switzerland
ming.liu@mavt.ethz.ch, btalper@sabanciuniv.edu, rsiegwart@ethz.ch

***Abstract*—Omnidirectional cameras are widely used for robotics applications in structured environments. However, because of the distorted field of view (FOV), it is hard to describe the primitive features extracted from them robustly. In this paper, we tackle the problem by using Histogram of Gradient (HoG) statistics for the regions of interest (ROI) in the neighbour of major vertical lines extracted from the panoramic image. As a validation, we compare the proposed algorithm with state-of-the-art based on two widely used datasets, leading to evidently better performance. We also introduce a scene reconstruction scenario using the proposed descriptor based on 1D Trifocal Tensor framework. The comparative results show the competence against other related works.**

I. Introduction

A. Motivation

Scene representation is a subtle problem, especially when non-standard imaging sensors such as omnidirectional camera is used. Although the calibration is less a problem using nowadays techniques [1], algorithms that independent from calibration result are still preferred, due to complexity and generalization potentials.

Omnidirectional camera is considered to be one of the most efficient sensors for environment modelling [2], [3]. However, a reliable descriptor for the conducted panoramic images is still required to be developed, moreimportant properly evaluated. Considering the characteristics of omnidirectional camera, we could see that the most reliable feature is the set of vertical lines perpendicular to the motion plane of the robot, since they are preserved regardless rotation and translation.

In this paper, we propose an adaptive descriptor for major vertical lines, which is inspired by and extended from [4]. We evaluate the performance in two steps. First we evaluate matching precision against [4] using two widely used datasets. Besides, we present a scene reconstruction scenario using trifocal tensor [5], as an application of the proposed feature.

B. Related Work

Several techniques are used to describe the surrounding environment of a robot. One of the major differences lies in the various descriptors used by structure reconstruction. We could see that many algorithms utilize keypoint based features on perspective cameras, e.g. PTAM [6] uses mainly FAST corners[7]; FAB-MAP [8] uses mainly SIFT [9] or SURF [10]. However, not many applications or descriptors have been reported on omnidirectional camera, such as the example depicted in figure 1. The main reason is the distortion introduced by the nonlinear transformation from the mirror shape. The nonuniform resolution will greatly affect the stableness of patch based descriptors.

This work was supported by the EU FP7 project NIFTi (contract # 247870).

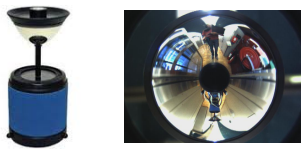


Fig. 1. Omnidirectional camera and panoramic image

There are two ways to represent the environment by images: First, descriptors can be extracted from the whole image, e.g. by Fourier transformation [11], [12], [13]. Among the existing algorithms designed for omnidirectional camera, “Fingerprint of places” [14] and FACT [15], [3] are color based features, where the vertical line is considered as an important hint for the formation of descriptors of the whole image. The second category is object oriented representations [16], [17], [18] namely feature based algorithms. Several lightweight keypoint descriptors were developed [19], [20] as well and got widely applied in scene recognition problems [21], [22].

We notice that beside the wide FOV, an important reason for choosing omnidirectional vision is that, when the camera is mounted perpendicularly to the plane of motion, the vertical lines of the scene are mapped into radial lines on the images. Regarding descriptor for such image primitives, [4] is defined for line description. However, we found it is hard to adapt it to robot translation, by which the length of critical vertical lines varies, due to the ROI is fixed even for completely different image frames. In this paper, we propose that the ROI need to be adaptable to different environment, and evaluate the parameter selection accordingly.

There are two groups of techniques to work on the vertical line matching. The first method deals with the individual line segments such as [23], [4] and the second one works with the grouping of the line segments [24], [25], [26]. Considering the complexity of the second group, in this paper we use the separation angle between two descriptor vectors as the primary metric to represent the similarity.

The panoramic images taken from omnidirectional camera can be used with the raw image or an unwrapped representa-

tion. In the case of full calibration, the raw image is usually taken as the algorithm input. However, when we focus on the vertical lines, the unwrapped image along the horizontal line is more feasible [27], [28]. This unwrapping process implies a calibration of the image center and extraction of the main circular shape as the right image in figure 1, which is dealt with by Hough Transformation described in section II.

In order to reconstruct scene appearances, the feature positions are to be recovered by geometrical constraints. In this paper, we use 1D Trifocal Tensor [29] to realize this reasoning process, which is mostly used in visual homing problem [29]. Comparing with other homing algorithms [30], [31], the trifocal tensor will result in not only robot positions, but also feature distributions. This provides a basis for scene reconstruction. We use the proposed features to provide a group of geometrical constraints in this work.

C. Arrangement

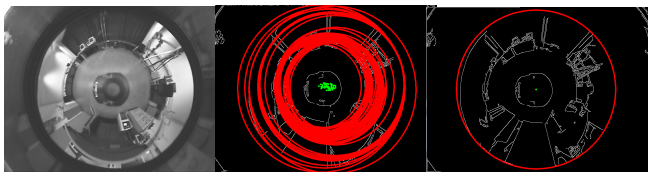
The rest of this paper is organized as follows. We first introduce the feature extraction and description in section II. Then, the scene reconstruction algorithm will be outlined in section III. The parameterization and evaluation will be carried out with widely cited datasets in section IV, followed by conclusion of this work in the end.

II. Proposed Descriptor

In this section, we introduce the major processes to detect salient features, namely vertical lines, and the descriptor formation.

A. Detection of Major Vertical Lines

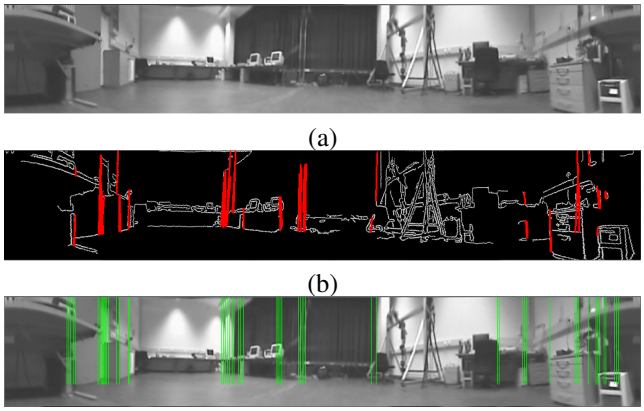
An unwrapped image will facilitate the extraction of major vertical lines, since all the radial lines are projected into vertical direction. Hough Circle Detection algorithm is first performed in order to obtain the radius of effective FOV and the center coordinate. The detection results is shown as figure 2. The outermost circle is taken as the effective FOV, since its inner part covers all valid information of the panoramic image. The estimated image center is taken by the circle center shown in figure 2(c).



(a) Raw Image (b) Detected circles (c) Outer circle

Fig. 2. All the circles detected in the raw image. The outermost circle is extracted.

The raw omnidirectional image is then unwrapped using interpolation as shown in figure 3(a) [1]. The unwrapping mapping makes the detection of vertical lines more straightforward. Using a 5-dimensional x-direction Sobel filter, all vertical lines are extracted. An instance is shown in figure 3(b). Based on the statistics of the accumulated strength of filtered results in x-direction, vertical lines with a length longer than average are considered as salient. (c) shows the detected major vertical lines projected into the raw unwrapped image.



(a)

(b)

(c)

Fig. 3. Major vertical lines extracted from an unwrapped panoramic image

B. Descriptor Formulation

In order to match the vertical lines across images, the formation of the descriptor is essential. Sometimes a tracking scheme can be adopted to help the matching process, where detected features have to be matched between two consecutive images [4]. In this work, we emphasize the appearance based matching without considering the tracking results.

We build the descriptor using the Histogram of Oriented Gradient (HoG). Considering the limitation of fixed circular shapes used by [4], we reshape the ROI by rectangles. For each major vertical line, a set of 6 ROI rectangles is extracted as shown in figure 4, where the width of rectangle can be adapted for different environments.

Fig. 4. The shape of the modified descriptor with varying $scaleX$. The width of the descriptor can change to adapt different environments.

In order to calculate the HoG efficiently, we first divide the orientation space ranged from $-\pi$ to π into N_b bins. Then two components of the image gradients for x- and y-directions, I_x and I_y, are calculated for each pixel in each rectangle. The counts per phase is then clustered, according to the discretized phase of the gradients Φ.

$$M = \sqrt{I_x^2 + I_y^2}, \quad \Phi = \arctan(I_y, I_x) \tag{1}$$

[1]The panoramic is with resolution 1024x176 in our tests.

Afterwards, the gradient magnitude M of each pixel is accumulated in the corresponding bin over the Φ space. An example of the calculated HoG is shown in figure 5.

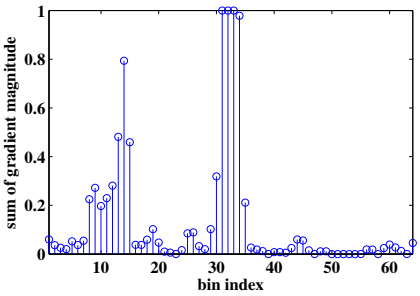


Fig. 5. An instance of non-normalized HoG with N_b = 32 bins.

The accumulated magnitude values are normalized in each rectange as the value with the maximum gradient magnitude is equal to one. All the bins with magnitude value greater than 0.1 (10% of the maximum value) are threshold as 0.1, then perform normalization again. This extra operation makes the descriptor more robust changes since the gradient magnitudes are more sensitive than orientation, in the case of illumination changes. At the end, three pairs of histograms H_1, H_2 and H_3 regarding left and right side of a vertical line are used as descriptor:

$$\begin{aligned} H_1 &= [H_{1,L}, H_{1,R}] \\ H_2 &= [H_{2,L}, H_{2,R}] \\ H_3 &= [H_{3,L}, H_{3,R}] \end{aligned} \tag{2}$$

We could see that two major parameters will determine the descriptor for a given image, i.e. number of bins for the HoG N_b and width of the rectangle, indicated by $scaleX$. For different specific environment, the optimal parameter set varies. The parameterization is evaluated in section IV.

C. Feature matching

In order to measure the similarity between two descriptors, we consider a descriptor as a vector with $6 \times N_b$ dimensions. Intuitively, we take the separate angle of two normalized descriptors $\mathbf{x}, \mathbf{y}$ as the measure of the distance, as:

$$\alpha(x, y) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{|\mathbf{x}||\mathbf{y}|} \tag{3}$$

where $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the inner product of the two descriptors. When comparing the features from two images A and B, letting $[A_1, A_2, .., A_m]$ be the descriptors of image A and $[B_1, B_2, .., B_n]$ be the descriptors of B, a positive matching is validated by the second best match is smaller than r_{th} ratio of the best match. An emperical r_{th} is 80%.

Considering the operation on a sequence of images, especially for tracking problems, we use a naive stratergy as follows. After matching the lines in the first two images, the same procedure is applied for the second and the third images for the vertical lines that have already been matched previously. As a sample result, a group of matched triple in three consecutive images is illustrated as in figure 6.

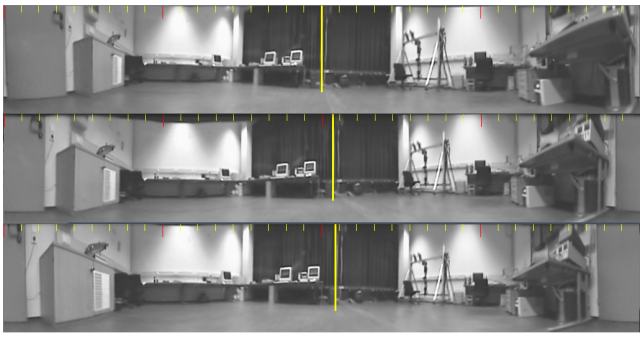


Fig. 6. An example of matched vertical line triples. The bearing information calculated for each image is used for trifocal tensor calculation.

III. Scene Reconstruction

Using trifocal tensor for scene reconstruction, the system needs bearing angles of matched features from three different robot positions. By using these three view bearing information, the 1D trifocal tensor can be calculated [29]. The trifocal tensor gives a constraint on relative position and orientation of three different robot positions. Given the estimated relative positions, by triangulating the landmarks, the geometrical structural information can be recovered.

A. Tensor Calculation

For the tensor calculation, we use the 1D trifocal tensor introduced in [32], [33] as basis. We concisely outline the process as follows.

The inputs of the tensor calculation process is at least seven bearing information triple that comes from three different robot positions. The bearing information from each major vertical line is kept in a state vector $u = (\sin\alpha, \cos\alpha)^T$, where α is the bearing angle of a line feature.

Following the notation of [32], θ's are used to present the robot heading and t_x, t_y are used to denote the translation in x- and y-direction for each local frame. The trifocal tensor is represented as:

$$\mathbf{T} = [T_{111}\ T_{112}\ T_{121}\ T_{122}\ T_{211}\ T_{212}\ T_{221}\ T_{222}]^T$$

where

$$\begin{aligned} T_{111} &= t_y^{'} sin(\theta^{''}) - t_y^{''} sin(\theta^{'}); \\ T_{112} &= t_y^{'} cos(\theta^{''}) + t_x^{''} sin(\theta^{'}); \\ T_{121} &= -t_x^{'} sin(\theta^{''}) - t_y^{''} cos(\theta^{'}); \\ T_{122} &= -t_x^{'} cos(\theta^{''}) + t_x^{''} cos(\theta^{'}); \\ T_{211} &= -t_y^{'} cos(\theta^{''}) + t_y^{''} cos(\theta^{'}); \\ T_{212} &= t_y^{'} sin(\theta^{''}) - t_x^{''} cos(\theta^{'}); \\ T_{221} &= t_x cos(\theta^{''}) - t_y^{''} sin(\theta^{'}); \\ T_{222} &= -t_x^{'} sin(\theta^{''}) + t_x^{''} sin(\theta^{'}). \end{aligned} \tag{4}$$

The trifocal constraints is rewritten using the coefficient matrix A and tensor T as (5).

$$\begin{aligned} A\mathbf{T} = [&u_1 u_1^{'} u_1^{''}\ u_1 u_1^{'} u_2^{''}\ u_1 u_2^{'} u_1^{''}\ u_1 u_2^{'} u_2^{''} \\ &u_2 u_1^{'} u_1^{''}\ u_2 u_1^{'} u_2^{''}\ u_2 u_2^{'} u_1^{''}\ u_2 u_2^{'} u_2^{''}]\mathbf{T} = 0 \end{aligned} \tag{5}$$

In order to solve approximated trifocal tensor T, the eigenvector associated with the smallest eigenvalue of the matrix A^TA is used, which theoretically obtained by singular value decomposition (SVD) of matrix A.

B. Solvers for scene reconstruction

Unlike the application of trifocal tensor in visual homing problem, the scene reconstruction problem greatly relies on the precision of the estimation of feature poses. The geometrical relations, such as translations and rotations, embedded in (4) are to be solved by minimizing equation 5. We try with three different solvers: Gauss-Newton algorithm, Levenberg-Marquardt algorithm and Stimulated Annealing. A typical

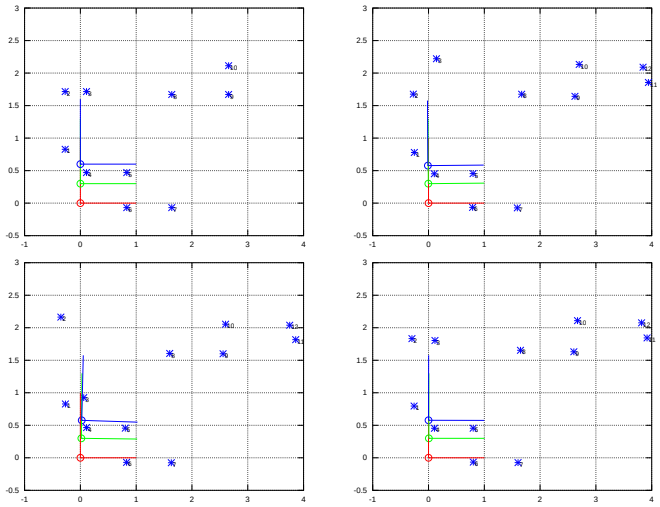


Fig. 7. Upper Left: Ground Truth, Upper Right: Gauss-Newton, Lower Left: Levenberg-Marquardt, Lower Right: Stimulated Annealing. The blue, green and red bars show the estimated robot position and orientation. The blue stars with numeric IDs indicate the reconstructed feature positions.

simulation result is shown as figure 7, where 1-degree Gaussian observation noise is introduced. Qualitatively, we can see that they lead to similar results in robot pose estimation [2]. However, due to the sensitiveness of the triangulation to rotation error, the feature reconstruction results vary much. Therefore, although stimulated annealing runs around 10 times slower than other two algorithms, as a global optimizer, we consider it as the primary solver in this work.

IV. Evaluation & Validation

A. Overview and Dataset

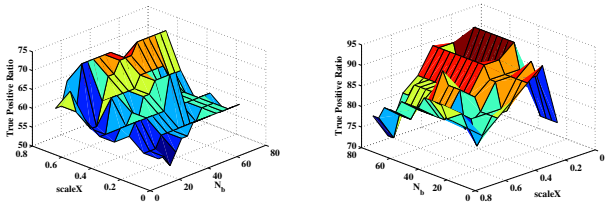
Two open source online datasets are adopted to validate the proposed descriptors. Both datasets are built with a mirrored omni-directional camera mounted on mobile wheeled robots for indoor environments [34], [35].

For each sample of a database, the feature matching is evaluated and compared with the state-of-art descriptor [4], in terms of true positive ratio. Please notice that the algorithm complexity for [4] and the proposed method is similar, since they both use HoG description. Therefore, the execution time is not taken for comparison.

[2]Due to space limit, we omit the quantitative results in this paper.

B. Parameter Selection

In order to optimize the parameters, for specific environments the two major parameters are to be selected based on sample statistics. The ranges of parameters are: N_b values are varied from 16 to 72, with increments by 4; the width of the descriptor($scaleX$) varies from 0.1 to 0.8, with increments by 0.1. We construct comparison matrices based on the true positive rates of the two descriptors on random samples. The results for the two datasets are shown in table I, II and table III, IV, respectively. For visualization purpose, we plot table I and III as shown in figure 8.



(a) COLD Dataset [35] (b) Vardy A1originalH Dataset [34]

Fig. 8. Visualization of evaluation tables for parameter selection.

We have the following observations for this part of evaluation:

- The proposed algorithm is evidently better performed in both datasets.
- The increment of number of bins for HoG will help the matching. However this emendation will have less effect when it reached to a certain large number. Considering the complexity of the histogram construction and feature matching is related to N_b, a "good enough" selection should be the knee value in the plot by figure 8.
- For both descriptors, the performance is worse for the COLD dataset. We observe the major reason is that the frame-edges of glass doors and windows in the COLD dataset triggers frequently wrong description by considering the scene behind. For the case of Vardy dataset, the appearances of the major vertical lines are usually not affected by perspective changes. This is the limitation for both descriptors.

As a result, the parameter set $\{N_b, scaleX\}$ for COLD dataset is $\{52, 0.6\}$, and $\{36, 0.4\}$ for Varday dataset. We see that the introduced adaptive parameter $scaleX$ greatly optimize the performance of the descriptor.

C. Reconstruction

We evaluate the performance for scene reconstruction by trifocal tensor. A typical failure case can be found in figure 7(b) and (c). The reconstructed robot positions are good enough for robot homing problems, however, the reconstruction of landmark distribution is poorly obtained. We found it is related to two characteristics of the coefficient matrix A, the smallest eigenvalue ($\min \lambda_i$) and the condition number of the matrix

$scaleX/N_b$	**16**	**20**	**24**	**28**	**32**	**36**	**40**	**44**	**48**	**52**	**56**	**60**	**64**	**68**	**72**
0.1	54.5	54.5	51.5	54.5	57.5	57.5	60.6	60.6	60.6	60.6	60.6	60.6	60.6	60.6	60.6
0.2	54.5	54.5	54.5	57.5	57.5	57.5	57.5	57.5	57.5	57.5	60.6	60.6	60.6	57.5	57.5
0.3	57.5	60.6	60.6	60.6	60.6	60.6	60.6	69.6	69.6	60.6	60.6	60.6	60.6	60.6	60.6
0.4	54.5	57.5	57.5	60.6	57.5	63.6	63.6	66.6	60.6	60.6	57.5	60.6	60.6	57.5	57.5
0.5	54.5	54.5	63.6	63.6	63.6	63.6	63.6	66.6	66.6	66.6	66.6	66.6	63.6	63.6	63.6
0.6	57.5	54.5	54.5	60.6	66.6	66.6	69.6	69.6	69.6	72.7	72.7	72.7	72.7	72.7	72.7
0.7	63.6	57.5	69.6	69.6	69.6	69.6	69.6	69.6	66.6	66.6	63.6	63.6	63.6	63.6	63.6
0.8	57.5	57.5	63.6	60.6	60.6	60.6	60.6	60.6	60.6	60.6	57.5	63.6	63.6	63.6	63.6

TABLE I

THE TRUE POSITIVE RATIO WITH THE PROPOSED DESCRIPTOR, BY VARYING N_b AND $scaleX$ (THE COLD DATASET).

N_b	**16**	**20**	**24**	**28**	**32**	**36**	**40**	**44**	**48**	**52**	**56**	**60**	**64**	**68**	**72**
	60.6	60.6	60.6	57.5	57.5	60.6	57.5	54.5	54.5	54.5	54.5	54.5	54.5	54.5	54.5

TABLE II

THE TRUE POSITIVE RATIO WITH [4], BY VARYING N_b (THE COLD DATASET).

$scaleX/N_b$	**16**	**20**	**24**	**28**	**32**	**36**	**40**	**44**	**48**	**52**	**56**	**60**	**64**	**68**	**72**
0.1	75.7	75.7	75.7	75.7	81.8	81.8	81.8	81.8	81.8	81.8	81.8	81.8	78.7	78.7	81.8
0.2	90.9	87.8	90.9	87.8	87.8	93.9	93.9	93.9	93.9	93.9	93.9	93.9	93.9	90.9	90.9
0.3	87.8	87.8	87.8	87.8	87.8	90.9	84.8	84.8	84.8	84.8	84.8	84.8	84.8	90.9	84.8
0.4	84.8	81.8	81.8	87.8	87.8	90.9	90.9	90.9	90.9	90.9	90.9	90.9	90.9	90.9	90.9
0.5	81.8	81.8	81.8	81.8	84.8	84.8	84.8	84.8	84.8	84.8	84.8	84.8	84.8	84.8	84.8
0.6	78.7	75.7	81.8	78.7	87.8	87.8	87.8	90.9	84.8	84.8	84.8	84.8	84.8	81.8	81.8
0.7	81.8	84.8	84.8	84.8	81.8	81.8	81.8	81.8	81.8	81.8	78.7	78.7	78.7	72.7	72.7
0.8	84.8	84.8	87.8	87.8	87.8	87.8	84.8	84.8	84.8	84.8	81.8	81.8	81.8	78.7	78.7

TABLE III

THE TRUE POSITIVE RATIO WITH THE PROPOSED DESCRIPTOR, BY VARYING N_b AND $scaleX$ (THE VARDY A1ORIGINALH DATASET)

N_b	**16**	**20**	**24**	**28**	**32**	**36**	**40**	**44**	**48**	**52**	**56**	**60**	**64**	**68**	**72**
	60.6	60.6	60.6	66.6	66.6	69.6	69.6	69.6	69.6	69.6	69.6	69.6	69.6	69.6	69.6

TABLE IV

THE TRUE POSITIVE RATIO WITH [4], BY VARYING N_b (THE VARDY A1ORIGINALH DATASET).

A ($cond(A)$). $\min \lambda_i$ defines the precision of trifocal tensor estimation by SVD, and $cond(A)$ reflect the stableness of the solution to equation 5. These two criteria can be taken as further assessment of the reconstruction quality.

1) Effect on the smallest eigenvalue: By increasing the standard deviation of the observation noise from 0.1 to 10, we show the uncertainty of the simulated features in figure 9, whereas $\min \lambda_i$ rises as depicted in figure 10. It implies that in order to have a reliable reconstruction, $\min \lambda_i$ needs to be as small as possible. Over a given threshold, the reconstruction results need to be discarded.

2) Effect on the conditional number: For the second characteristic $cond(A)$, the relation to the variety in the bearing information is investigated. A larger conditional number will in general lead to unreliable solutions for linear systems. In order to test how the perspective differences affect the robustness of reconstruction, we use different distances among the observing poses, depicted in figure 11. Intuitively, we can imagine that the closer the robot positions are, the more confused for the scene recognition. Figure 12 validates this assumption by plotting the relation between $cond(A)$ and the mean distance between two observation poses. We can observe that a larger distance will optimize the quality of the reconstruction, but it usually leads to less positive matches for real data. Therefore compromise is required for threshold selection. [3]

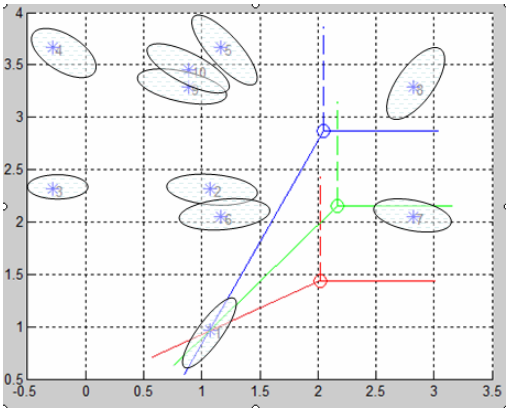


Fig. 9. Landmark locations with uncertainty.

3) Reconstruction result: Given the analysis on parameterization and quality justification, the scene reconstruction is carried out by firstly thresholding the aforementioned criteria.

[3]In this work, threshold for $\min \lambda_i$ is 0.02, and threshold for $cond(A)$ is 100.

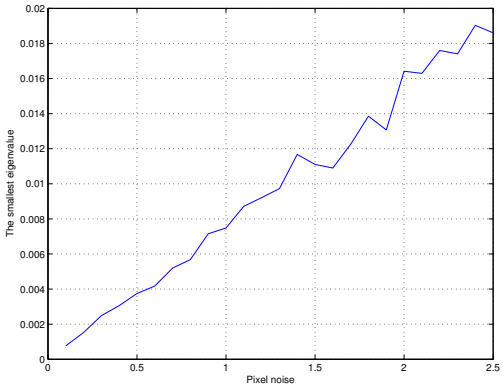


Fig. 10. Pixel noise vs The smallest eigenvalue.

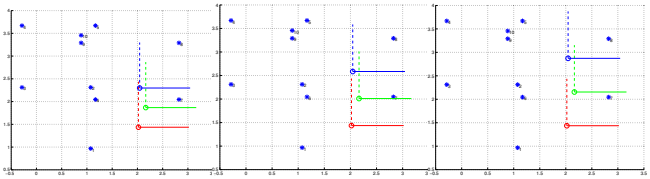


Fig. 11. Examples of various distances among robot positions. The effect of the distance variance is investigated while keeping the same feature distribution.

Unreliable matched feature sets are discarded. Then, geometrical information is calculated from equation (4) and (5) using Stimulated Annealing, using a single-shot odometry measure to correct the transformation scale between image space and real world. A qualitative result is shown in figure 13 using the images in figure 6.

V. Conclusion

In this paper, we first introduced an adaptive descriptor designed for omnidirectional camera. It works on the panoramic images, independent of intrinsic calibration. It outperforms the state-of-the-art, in terms of recall precision as well. The proposed descriptor is validated by a scene reconstruction scenario. Beside, two criteria for scene recognition problem are proposed and validated through simulation. As future work,

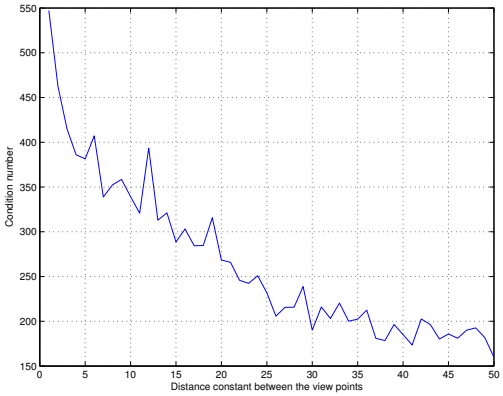


Fig. 12. Distance vs Condition number.

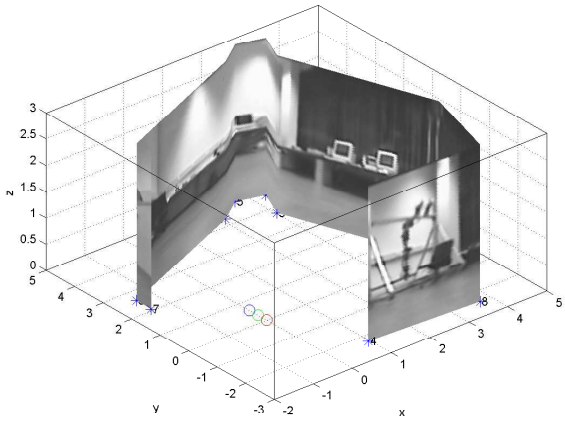


Fig. 13. Reconstructed environment in 3D by trifocal tensor

we will focus on applications using the proposed descriptor and quantitative assessment of the reconstruction quality.

References

[1] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006.

[2] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, "Omni-directional vision for robot navigation," in *Proceedings of the IEEE Workshop on Omnidirectional Vision*. IEEE, 2000, pp. 21–28.

[3] M. Liu and R. Siegwart, "Dp-fact: Towards topological mapping and scene recognition with color for omnidirectional camera," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, may 2012, pp. 3503 –3508.

[4] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A robust descriptor for tracking vertical lines in omnidirectional images and its use in mobile robotics," *International Journal of Robotics Research*, 2009, special Issue on Field and Service Robotics.

[5] H. Becerra, G. López-Nicolas, and C. Sagüés, "Omnidirectional visual control of mobile robots based on the 1d trifocal tensor," *Robotics and Autonomous Systems*, vol. 58, no. 6, pp. 796–808, 2010.

[6] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, Nara, Japan, November 2007.

[7] M. Trajković and M. Hedley, "Fast corner detection," *Image and Vision Computing*, vol. 16, no. 2, pp. 75–87, 1998.

[8] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[9] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Lecture notes in computer science*, vol. 3951, p. 404, 2006.

[11] X. Meng, Z. Wang, and L. Wu, "Building global image features for scene recognition," *Pattern Recognition*, 2011.

[12] A. Pretto, E. Menegatti, Y. Jitsukawa, R. Ueda, and T. Arai, "Image similarity based on discrete wavelet transform for robots with low-computational resources," *Robotics and Autonomous Systems*, vol. 58, no. 7, pp. 879–888, 2010.

[13] L. Payá, L. Fernández, A. Gil, and O. Reinoso, "Map building and monte carlo localization using global appearance of omnidirectional images," *Sensors*, vol. 10, no. 12, pp. 11 468–11 497, 2010.

[14] P. Lamon, A. Tapus, E. Glauser, N. Tomatis, and R. Siegwart, "Environmental modeling with fingerprint sequences for topological global localization," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 4, 2003.

[15] M. Liu, D. Scaramuzza, C. Pradalier, R. Siegwart, and Q. Chen, "Scene recognition with omnidirectional vision for topological map using lightweight adaptive descriptors," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, oct. 2009, pp. 116 –121.

[16] A. Bosch, A. Zisserman, and X. Munoz, "Scene classification via plsa," *European Conference on Computer Vision (ECCV)*, pp. 517–530, 2006.

[17] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context-based vision system for place and object recognition," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2003.

[18] S. Vasudevan, S. Gachter, V. Nguyen, and R. Siegwart, "Cognitive maps for mobile robots–an object based approach," *Robotics and Autonomous Systems*, vol. 55, no. 5, pp. 359 – 371, 2007, from Sensors to Human Spatial Concepts. [Online]. Available: http://www.sciencedirect.com/science/article/B6V16-4MY0MK7-1/2/e379fd59a33b6d0a42355ba120c444e9

[19] J. Wu and J. Rehg, "Centrist: A visual descriptor for scene categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1489–1501, 2010.

[20] M. Calonder, V. Lepetit, and P. Fua, "Keypoint signatures for fast learning and recognition," *European Conference on Computer Vision (ECCV)*, pp. 58–71, 2008.

[21] J. Wu, H. Christensen, and J. Rehg, "Visual place categorization: problem, dataset, and algorithm," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2009, pp. 4763–4770.

[22] A. Ranganathan, "PLISS: Detecting and Labeling Places Using Online Change-Point Detection," *Proceedings of Robotics: Science and Systems, Zaragoza, Spain*, 2010.

[23] Z. Zhang, "Token tracking in a cluttered scene," *Image and Vision Computing*, vol. 12, no. 2, pp. 110–120, 1994.

[24] J. Crowley, P. Stelmaszyk, T. Skordas, and P. Puget, "Measurement and integration of 3-d structures by tracking edge lines," *International Journal of Computer Vision*, vol. 8, no. 1, pp. 29–52, 1992.

[25] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the hausdorff distance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 9, pp. 850–863, 1993.

[26] R. Deriche and O. Faugeras, "Tracking line segments," *Image and Vision Computing*, vol. 8, no. 4, pp. 261–270, 1990.

[27] S. Nayar, "Catadioptric omnidirectional camera," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE, 1997, pp. 482–488.

[28] T. Mauthner, F. Fraundorfer, and H. Bischof, "Region matching for omnidirectional images using virtual camera planes," in *Proc. of Computer Vision Winter Workshop*, 2006.

[29] M. Aranda, G. Lopez-Nicolas, and C. Sagues, "Omnidirectional visual homing using the 1D trifocal tensor," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 2444–2450.

[30] M. Liu, C. Pradalier, F. Pomerleau, and R. Siegwart, "The role of homing in visual topological navigation," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, oct. 2012, pp. 567 –572.

[31] M. Liu, C. Pradalier, Q. Chen, and R. Siegwart, "A bearing-only 2d/3d-homing method under a visual servoing framework," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, may 2010, pp. 4062 –4067.

[32] J. Guerrero, A. Murillo, and C. Sagues, "Localization and matching using the planar trifocal tensor with bearing-only data," *Robotics, IEEE Transactions on*, vol. 24, no. 2, pp. 494–501, 2008.

[33] O. Faugeras, L. Quan, and P. Sturm, "Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera," *European Conference on Computer Vision (ECCV)*, pp. 36–52, 1998.

[34] A. Vardy, "Panoramic image database." [Online]. Available: http://www.ti.uni-bielefeld.de/html/research/avardy/index.html

[35] A. Pronobis and B. Caputo, "COLD: The CoSy localization database," *The International Journal of Robotics Research*, vol. 28, no. 5, pp. 588–594, 2009.