Closing the Loop in Appearance-Guided Structure-from-Motion for Omnidirectional Cameras

Davide Scaramuzza, Friedrich Fraundorfer, Marc Pollefeys, Roland Siegwart

▶ **To cite this version:**

Davide Scaramuzza, Friedrich Fraundorfer, Marc Pollefeys, Roland Siegwart. Closing the Loop in Appearance-Guided Structure-from-Motion for Omnidirectional Cameras. The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS, Rahul Swaminathan and Vincenzo Caglioti and Antonis Argyros, Oct 2008, Marseille, France. inria-00325324

HAL Id: inria-00325324

https://inria.hal.science/inria-00325324

Submitted on 28 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author manuscript, published in "Dans The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS (2008)"

inria-00325324, version 1 - 28 Sep 2008

Closing the Loop in Appearance-Guided Structure-from-Motion for Omnidirectional Cameras

Davide Scaramuzza[1], Friedrich Fraundorfer[2],
Marc Pollefeys[2], and Roland Siegwart[1]

[1]Autonomous Systems Lab, ETH Zurich
[2]Computer Vision and Geometry Group, ETH Zurich

Abstract. In this paper, we present a method that allows us to recover a 400 meter trajectory purely from monocular omnidirectional images very accurately. The method uses a novel combination of appearance-guided structure from motion and loop closing. The appearance-guided monocular structure-from-motion scheme is used for initial motion estimation. Appearance information is used to correct the rotation estimates computed from feature points only. A place recognition scheme is employed for loop detection, which works with a visual word based approach. Loop closing is done by bundle adjustment minimizing the reprojection error of feature matches. The proposed method is successfully demonstrated on videos from an automotive platform. The experiments show that the use of appearance information leads to superior motion estimates compared to a purely feature based approach. And we demonstrate a working loop closing method which eliminates the residual drift errors of the motion estimation. Note that the recovered trajectory is one of the longest ones ever reported with a single omnidirectional camera.

1 Introduction

Robust and reliable trajectory recovery for automotive applications using visual input only needs a very accurate motion estimation step and loop closing for removing the inevitably accumulated drift. The first focus of this paper is in using the appearance information to improve the results of feature based motion estimation. The second focus is in removing accumulated drift with loop closing.

In this paper, we use a single calibrated catadioptric camera mounted on the roof of the car (Fig. 1). We assume that the vehicle undergoes a purely two-dimensional motion over a predominant flat ground. Furthermore, because we want to perform motion estimation in city streets, flat terrains, as in well as in motorways where buildings or 3D structure are not always present, we estimate the motion of the vehicle by tracking the ground plane.

The first step of our approach is to extract SIFT keypoints [1] from the scene all around the car and match them between consecutive frames. After RANSAC based outlier removal [2], we use these features to compute the translation in

Fig. 1. Our vehicle with the omnidirectional camera (blue circle). The field of view is indicated by the red lines.

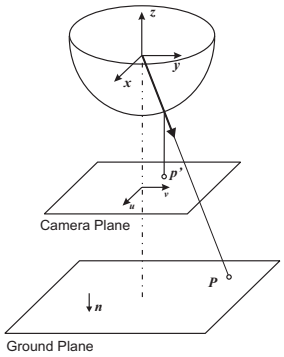


Fig. 2. Our omnidirectional camera model.

the heading direction only. To estimate the rotation angle of the vehicle we instead use an appearance based method. We show that by using appearance information our result outperforms the pure feature based approach. At the same time as motion estimation, a loop detection algorithm is running. We use a visual word based approach [3, 4] that is very fast and highly scalable. In addition, we designed a geometric loop verification especially for omnidirectional images. Loop closing is then finally done in an optimization step using bundle adjustment. The method is demonstrated on motion estimation. We demonstrate this method on video data from a 400m trajectory and show that the initial motion estimation is already very accurate.

The remainder of the paper is organized as follows. Section 2 reviews the related work. Section 3 describes our homography based motion estimation which is used for translation estimation. Section 4 describes the details about the appearance guided Structure from Motion (SfM) which corrects the rotation estimates. Section 5 details the steps of the whole SfM algorithm. Section 6 describes the loop closing algorithm. Finally, Section 7 is dedicated to the experimental results.

2 Related Work

Structure from motion and motion estimation (also called visual odometry) with omnidirectional cameras has already been investigated from various groups [5–7]. The benefit of camera trajectory determination using large field of view was firstly demonstrated by Svoboda *et al.* [8] and further recognized by Chang and Hebert [9]. However, in those works SfM was performed only over short distances (up to a few meters). Conversely, in this paper we concentrate on SfM and accurate trajectory recovery over long distances (hundreds of meters).

Motion estimation with omnidirectional cameras for long trajectories has been investigated by [10–12]. In [10], Corke *et al.* provided two approaches for

inria-00325324, version 1 - 28 Sep 2008

inria-00325324, version 1 - 28 Sep 2008

monocular visual odometry based on omnidirectional imagery. As their approach was conceived for a planetary rover, they performed experiments in the desert and therefore used keypoints from the ground plane. In the first approach, they used optical flow computation with planar motion assumption while in the second one SfM with no constrained motion using an extended Kalman filter. The optical flow based approach gave the best performance over 250 meters but the trajectory was not accurately recovered showing a large drift of the rotation estimation. Another approach with robust and reliable motion estimation was presented by Lhuillier [11] where only keypoint tracking and bundle adjustment were used to recover both the motion and the 3D map. None of these methods however address loop detection and loop closing.

Loop closing in general was described by Bosse *et al.* [13]. The approach worked by map matching of laser scans using 3D points and 3D lines as features. Map matching with 3D point and 3D line features from image data was demonstrated in [14]. Experiments were shown on indoor and urban areas with plenty of line features. No result is shown on image data similar to ours, where almost no line features are present. Loop detection in a similar manner to ours is also described in [15]. In their case a loop detection probability is computed from the visual similarity of images. In contrast to our approach no geometric verification of visual features is performed and no actual loop closure using the detected loops is done. Loop detection and loop closure is successfully demonstrated in [16] where they use a laser range finder to get the initial trajectory and image data for loop detection. Loop closure however also relies on information from the laser range finder, whereas our proposed approach uses visual features only.

In this paper we extend our previous work on appearance-guided visual odometry for outdoor ground vehicles [17, 18] by addressing the problem of loop detection and closing. To make this paper self consistent, we summarize our previous work in Section 3 and 4.

3 Homography Based Motion Estimation

Our initial motion estimation proceeds along the lines of our previous work (please refer to [17, 18] for major details). The method uses planar constraints and point tracking to compute the motion parameters. As we assume planar motion and that the camera is orthogonal to the ground plane with quite a good approximation, only two points are needed to estimate the motion parameters up to a scale (the scale is then recovered from the height of the camera). Then, after a two-point RANSAC based outlier removal, rotation and translation parameters between consecutive frames are computed from the remained inliers. For this the homography decomposition of Triggs [19] is used being adapted to omnidirectional images. A subsequent non-linear refinement improves the accuracy. In this we constrain the minimization so that the rotation is about the plane normal and the translation is parallel to the same plane as we assume planar motion.

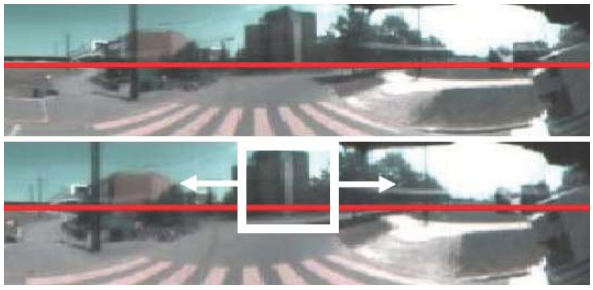


Fig. 3. Two unwrapped omnidirectional images taken at consecutive time stamps. For reasons of space, here only one half of the whole 360 *deg* is shown. The red line indicates the horizon.

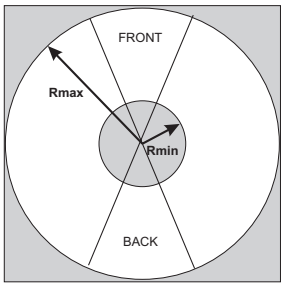


Fig. 4. The cylindrical panorama is obtained by unwrapping the white region.

4 Visual Compass

Unfortunately, when using point features to estimate the motion, the resulting rotation is extremely sensitive to systematic errors due to the intrinsic calibration of the camera or the extrinsic calibration between the camera and the ground plane. This effect is even more accentuated with omnidirectional cameras due to the large distortion introduced by the mirror. In addition to this, integrating rotational information over the time has the major drawback of generally becoming less and less accurate as integration introduces additive errors at each step. An example of camera trajectory recovered using only the feature based approach described in Section 3 is depicted in Fig. 5 (blue trajectory).

To improve the accuracy of the rotation estimation, we use an appearance based approach. This approach was inspired by the work of Labrosse [20], which describes a method to use omnidirectional cameras as visual compass.

Directly using the appearance of the world as opposed to extracting features or structure of the world is attractive because methods can be devised that do not need precise calibration steps. Here, we describe how we implemented our visual compass.

For ease of processing, every omnidirectional image is unwrapped into cylindrical panoramas (Fig. 3). The unwrapping considers only the white region of the omnidirectional image that is depicted in Fig 4. We call these unwrapped versions "appearances". If the camera is perfectly vertical to the ground, then a pure rotation about its vertical axis will result in a simple column-wise shift of the appearance in the opposite direction. The exact rotation angle could then be retrieved by simply finding the best match between a reference image (before rotation) and a column-wise shift of the successive image (after rotation). The best shift is directly related to the rotation angle undertaken by the camera. In the general motion, translational information is also present. This general case will be discussed later.

The input to our rotation estimation scheme is thus made of appearances that need to be compared. To compare them, we use the Euclidean distance. The

inria-00325324, version 1 - 28 Sep 2008

inria-00325324, version 1 - 28 Sep 2008

Euclidean distance between two appearances I_i and I_j, with I_j being column-wise shifted (with column wrapping) by α pixels, is:

$$d(I_i, I_j, \alpha) = \sqrt{\sum_{k=1}^{h} \sum_{h=1}^{w} \sum_{l=1}^{c} |I_i(k,h,l) - I_j(k,h-\alpha,l)|^2} \tag{1}$$

where $h \times w$ is the image size, and c is the number of color components. In our experiments, we used the RGB color space, thus having three color components per pixel.

If α_m is the best shift that minimizes the distance between two appearances I_i and I_j, the rotation angle $\Delta\vartheta$ (in degrees) between I_i and I_j can be computed as:

$$\Delta\vartheta = \alpha_m \cdot \frac{360}{w} \tag{2}$$

The width w of the appearance is the width of the omnidirectional image after unwrapping and can be chosen arbitrarily. In our experiments, we used $w = 360$, that means the angular resolution was 1 pixel per degree. To increase the resolution to 0.1 *deg*, we used cubic spline interpolation with 0.1 pixel precision. We also tried larger image widths but we did not get any remarkable improvement in the final result. Thus, we used $w = 360$ as the unwrapping can be done in a negligible amount of time.

The distance minimization in (1) makes sense only when the camera undergoes a pure rotation about its vertical axis, as a rotation corresponds to a horizontal shift in the appearance. In the real case, the vehicle is moving and translational component is present. However, the "pure rotation" assumption still holds if the camera undergoes small displacements or the distance to the objects (buildings, tree, etc.) is big compared to the displacement. In the other cases, this assumption does not hold for the whole image but an improvement that can be done over the theoretical method is to only consider parts of the images, namely the front and back part (Fig. 4). Indeed, the contribution to the optical flow by the motion of the camera is not homogeneous in omnidirectional images; a forward/backward translation mostly contributes in the regions corresponding to the sides of the camera and very little in the parts corresponding to the front and back of the camera, while the rotation contributes equally everywhere.

Because we are interested in extracting the rotation information, only considering the regions of the images corresponding to the front and back of the camera allows us to reduce most of the problems introduced by the translation, in particular sudden changes in appearance (parallax).

According to the last considerations, in our experiments we use a reduced Field Of View (FOV) around the front and back of the camera (Fig. 4). A reduced field of view of about 30 *deg* around the front part is shown by the white window in Fig. 3. Observe that, besides reducing the FOV of the camera in the horizontal plane, we operate a reduction of the FOV also in the vertical plane, in particular

inria-00325324, version 1 - 28 Sep 2008

under the horizon line. The objective is to reduce the influence of the changes in appearance of the road. The resulting vertical FOV is 50 *deg* above and 10 *deg* below the horizon line (the horizon line is indicated in red in Fig. 3).

5 Motion Estimation Algorithm

As we already mentioned, the appearance based approach provides rotation angle estimates that are more reliable and stable than those output by the pure feature based approach. Here, we describe how we combined the rotation angle estimates of Section 4 with the camera translation estimates of Section 3.

In our experiments, the speed of the vehicle ranged between 10 and 20 Km/h while the images were constantly captured at 10 Hz. This means that the distance covered between two consecutive frames ranged between 0.3 and 0.6 meters. For this short distance, the kinematic model of the camera configuration (x, y, θ), which contains its 2D position (x,y) and orientation θ, can be approximated in this way:

$$\begin{cases} x_{i+1} = x_i \ + \ \delta\rho_i \cos(\theta_i + \frac{\delta\theta_i}{2}) \\ y_{i+1} = y_i \ + \ \delta\rho_i \sin(\theta_i + \frac{\delta\theta_i}{2}) \\ \theta_{i+1} = \theta_i \ + \ \delta\theta_i \end{cases} \tag{3}$$

where we use $\delta\rho = |\mathbf{T}|\ h$ and $\delta\theta = \Delta\vartheta$. $|\mathbf{T}|$ is the length of the translation vector assuming the camera at unit distance from the ground plane; h is the scale factor (i.e. in our case this is the height of the camera to the ground plane). The camera rotation angle $\Delta\vartheta$ is computed as in (2). Observe that we did not use at all the rotation estimates provided by the feature based method of Section 3.

Now, let us resume the steps of our motion estimation scheme, which have been described in Sections 3 and 4. Our omnidirectional visual odometry operates as follows:

1. Acquire two consecutive frames. Consider only the region of the omnidirectional image, which is between *Rmin* and *Rmax* (Fig. 4).
2. Extract and match SIFT features between the two frames. Use the double consistency check to reduce the number of outliers. Then, use the calibrated omnidirectional camera model to normalize the feature coordinates so that the z-coordinate is equal to -1 (see Fig. 2).
3. Use 4-point RANSAC to reject points that are not coplanar (at lest 4 points are needed to compute a homography).
4. Apply the Triggs algorithm followed by non-linear refinement described in Section 3 to estimate $\mathbf{R}$ and $\mathbf{T}$ from the remaining inliers.
5. Unwrap the two images and compare them using the appearance method described in Section 4. In particular, minimize (1), with reduced field of view, to compute the column-wise shift α_m between the appearances and use (2) to compute the rotation angle $\Delta\vartheta$.
6. Use $\delta\rho = |\mathbf{T}|\ h$ and $\delta\theta = \Delta\vartheta$ and integrate the motion using (3).
7. Repeat from step 1.

inria-00325324, version 1 - 28 Sep 2008

6 Vision based loop detection and closing

Loop closing is essential to remove accumulated drift errors in the camera trajectory. Our loop closing proceeds along three steps. The first step is loop detection which is done by a visual place recognition scheme. Next geometric correspondences between the two matching places are established. Finally loop closing is performed by optimization of structure and motion using bundle adjustment.

6.1 Loop detection by visual place recognition

For loop detection we use a visual word based place recognition system. Each image that got acquired is assigned a place in the world. To recognize places that have been visited before, image similarity (based on local features) is used. Our approach is along the lines of the method described in [3]. Firstly local features are extracted from images. We use Difference of Gaussian (DoG) keypoints and compute a SIFT feature vector for each keypoint. Each SIFT feature vector is quantized by a hierarchical vocabulary tree. All visual words from one image form a document vector which is a v-dimensional vector where v is the number of possible visual words. It is usually extremely sparse. For place recognition the similarity between the query document vector to all document vectors in a database is computed. As similarity score we use the L_2 distance between document vectors. The organization of the database as an inverted file and the sparseness of the document vectors allows a very efficient scoring. For scoring, the different visual words are weighted based on the Inverse Document Frequency (IDF) measure. Place recognition in our case works as an online approach. For each new image the similarity to all the images in the database is computed. The n top ranked images are stored as loop hypotheses. The current image is then stored in the database as well. The loop hypotheses are then geometrically verified. If one hypothesis passes this verification the loop closing optimization will be invoked. Our place recognition is very fast and maintains online speed for very large databases (almost up to 1 million images). The visual word based image representation is extremely compact and allows the storage of many places.

6.2 Geometric correspondences and loop verification

For loop closing geometric correspondences need to be established for the loop hypothesis. For our loop closing we need to get point matches between the images from matching places. These are created from the geometric verification step. Our geometric verification is designed for our omnidirectional images. Besides the image similarity our place recognition also returns point correspondences between the two images. These point correspondences are now used for geometric verification. For verification we first compute a polar coordinate representation (r, ϕ) from the (x, y) image coordinates. We assume that both images only differ by an in-plane rotation, neglecting a possible translation. Then we rotate the features of one of the images to search for the rotation that produces the maximum

feature overlap. If the number of overlapping features is higher than a threshold S the loop hypothesis is counted as correct. An inlier set from overlapping features is then used in the loop closing process.

6.3 Loop closing

Loop closing is carried out as bundle adjustment [21] over camera poses and structure (3D points). In our case poses and points are restricted to the ground plane, so that we have fewer parameters than standard bundle adjustment. Initial poses and 3D points come from the SfM algorithm. The detected loop now adds an extra constraint to the optimization. From loop detection we know the two poses p_0 and p_1 so far, as well as 2D image matches between the images I_0 and I_1. As we assume planar motion and with known height of the camera we can compute 3D points directly from 2D image points. We do this for image I_0 with pose p_0 creating $\mathbf{X_0}$. Reprojecting $\mathbf{X_0}$ into image I_1 should ideally create points that exactly lie on the original features in I_1 if pose p_1 is correct. We denote them p'_0. In loop closing this reprojection error is now minimized by changing the initial camera poses and structure. Loop closing usually cannot be done with online speed, but it is scalable up to large maps using for instance the method described in [22].

7 Results

7.1 Motion estimation

This experiment shows motion estimation results using the proposed approach and also compares it to a feature only based approach. It was tested with data from a real vehicle equipped with a central omnidirectional camera. A picture of our vehicle (a Smart) is shown in Fig 1. Our omnidirectional camera, composed of a hyperbolic mirror (KAIDAN 360 One VR) and a digital color camera (SONY XCD-SX910, image size 640×480 pixels), was installed on the front part of the roof of the vehicle. The frames were grabbed at 10 Hz and the vehicle speed ranged between 10 and 20 Km/h. The resulting path estimated by our SfM algorithm using a horizontal reduced FOV of 10 *deg* is shown in figures 5, 6, and 7. The results of the feature only based approach in Fig. 5 are shown in blue. The trajectory with our appearance based method is shown in red. From the aerial view in 6 it is clear that our method produces the correct trajectory.

In this experiment, the vehicle was driven along a 400 meter long loop and returned to its starting position (pointed to by the yellow arrow in Fig. 6). The estimated path is indicated with red dots in Fig. 6 and is shown superimposed on the aerial image for comparison. The final error at the loop closure is about 6.5 meters. This error is due to the unavoidable visual odometry drift; however, observe that the trajectory is very well estimated until the third 90-degree turn. After the third turn, the estimated path deviates smoothly from the expected path instead of continuing straight. After road inspection, we found that the

inria-00325324, version 1 - 28 Sep 2008

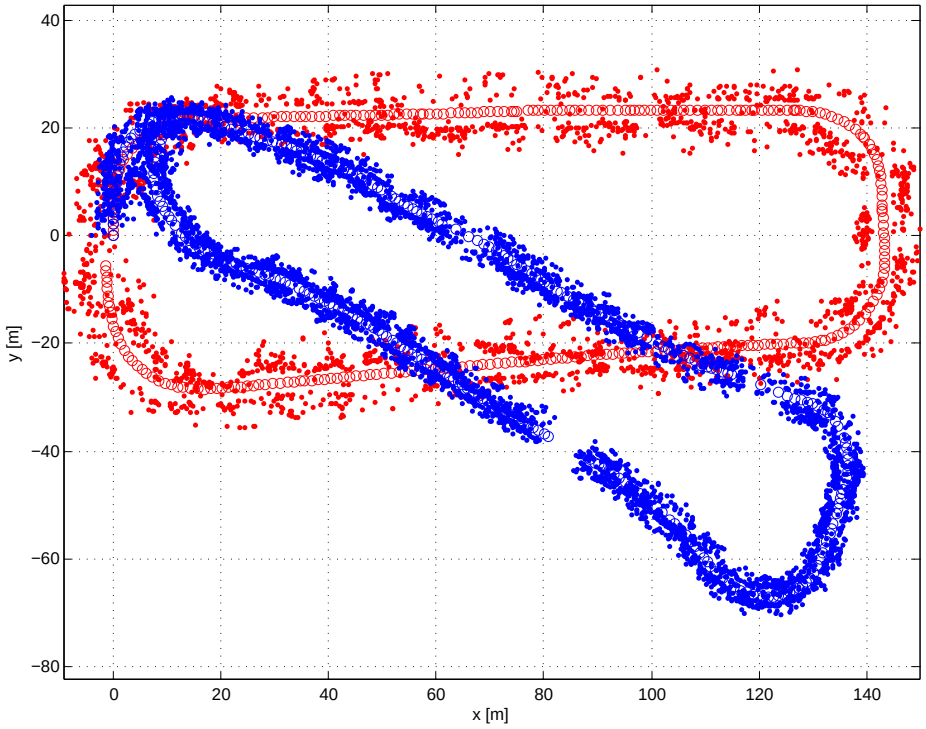


Fig. 5. Comparison between the standard feature based approach (blue) and the approach combining features with visual compass proposed in this paper (red).Circles are camera poses, dots are reconstructed feature points.

reason for this were most likely three 0.3 meter tall road humps (pointed to by the cyan arrow in Fig. 5).

The content of Fig. 7 is very important as it allows us to evaluate the quality of motion estimation. In this figure, we show a textured top viewed 2D reconstruction of the whole path. Observe that this image is not an aerial image but is an image mosaicing. Every input image of this mosaic was obtained by an Inverse Perspective Mapping (IPM) of the original omnidirectional image onto an horizontal plane. After being undistorted through IPM, these images have been merged together using the 2D poses estimated by our visual odometry algorithm. The estimated trajectory of the camera is shown superimposed with red dots. If the reader visually compares the mosaic (Fig. 7) with the corresponding aerial image (Fig. 6), he will recognize in the mosaic the same elements that are present in the aerial image, that is, trees, white footpaths, pedestrian crossings, roads' placement, etc. Furthermore, as can be verified, the position of these elements in the mosaic fits well the position of the same elements in the aerial image.

inria-00325324, version 1 - 28 Sep 2008

Fig. 6. The estimated path (before loop closing) superimposed onto a Google Earth image of the test environment. The scale is shown at the lower left corner.

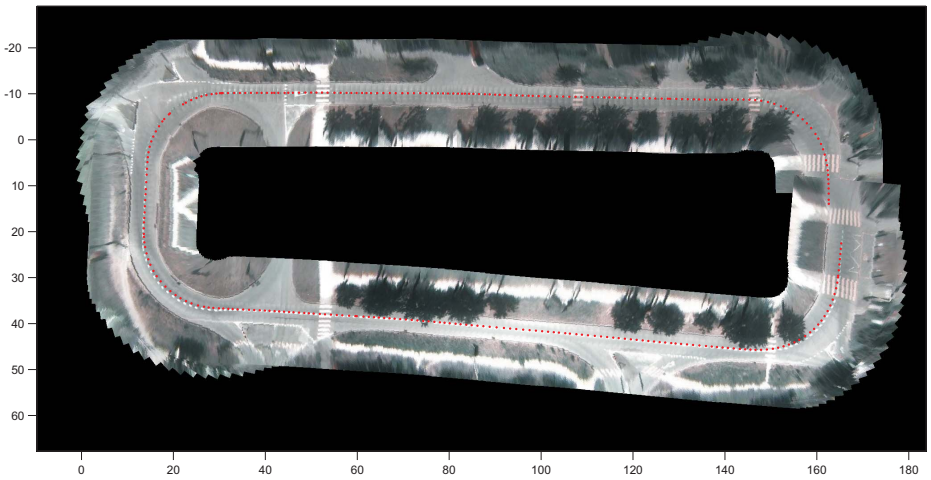


Fig. 7. Image mosaicing that shows a textured 2D reconstruction of the estimated path before closing the loop.

7.2 Motion estimation with loop closing

We performed the same experiment as in the previous section with loop closing running in parallel. The inlier threshold S (to accept a loop hypothesis) was set to 10, which led to the detection of a loop between frame 2 and 298. 10 geometric matches were extracted and used as additional constraints in the bundle adjustment. Fig. 8 shows the results after loop closing. The optimized trajectory is shown in red, while the initial parameters are shown in blue and black. The

inria-00325324, version 1 - 28 Sep 2008

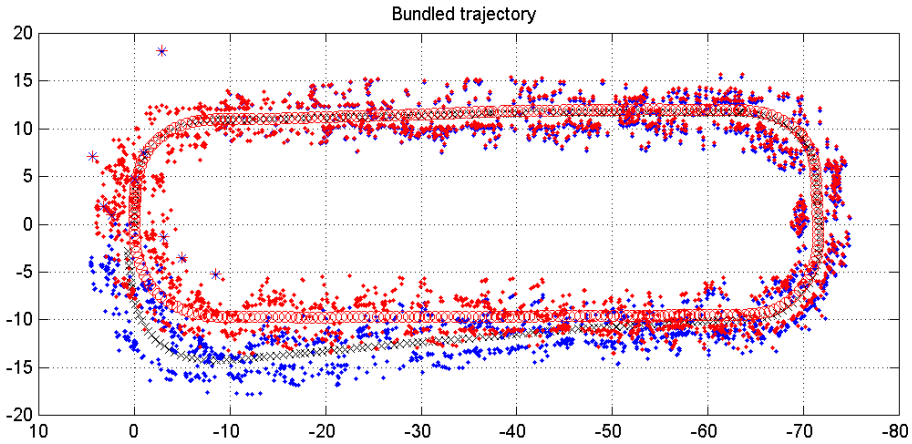


Fig. 8. Structure and motion after loop closing (red). Initial estimates are shown in blue and black. The loop is nicely closed.

circles are the camera position and the dots are reconstructed feature points. The loop got nicely closed. Fig. 9 shows the path after loop closing superimposed on a Google Earth aerial view. It is an almost perfect fit. Fig. 10 shows the image mosaic created from the optimized path. Fig. 11 visualizes the loop constraints, the feature points p_0 and p'_0. Before loop closing they don't coincide, the distance between them is high. Loop closing will now try to minimize the distance. As it is visible in the right image this could be done successfully, the total distance of 851.8 pixel could be reduced to 13.8 pixel.

8 Conclusion

It this paper, we presented a reliable and robust method for structure from motion for omnidirectional cameras. Initial motion estimates from feature tracks were corrected using appearance information. Finally loop detection and loop closing removed accumulated drift. The proposed method runs fast and is scalable. The method however assumes planar motion which is in particular the case in many automotive and robotics applications. Only features from the ground plane are tracked, i.e. this approach will also work outside urban areas and it is perfectly suited for outdoor areas. In our experiments we demonstrated motion estimation on a challenging 400m long path that is one of the longest distances ever reported with a single omnidirectional camera. We showed that the use of appearance information enormously improved the motion estimates resulting in a very accurate estimate. Loop closing finally again improved the accuracy by removing accumulated drift.

inria-00325324, version 1 - 28 Sep 2008

Fig. 9. The estimated path after loop closing superimposed onto a Google Earth.

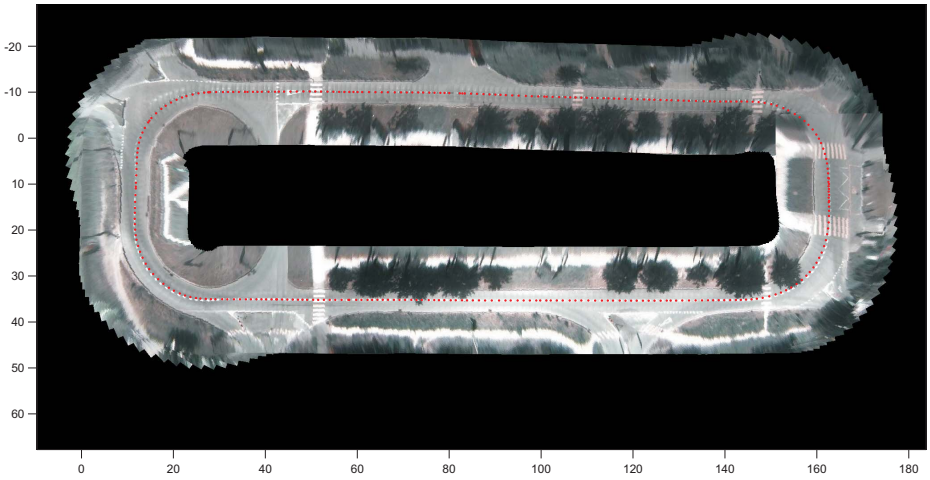
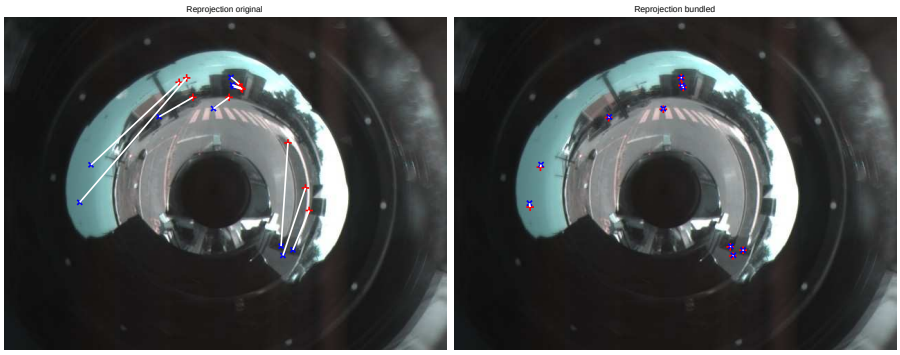


Fig. 10. Image mosaicing of the estimated path after loop closing.

9 Acknowledgment

The research leading to these results has received funding from the European Commission Division FP6-IST Future and Emerging Technologies under the contract FP6-IST-027140 (BACS: Bayesian Approach to Cognitive Systems).

inria-00325324, version 1 - 28 Sep 2008



(a) (b)

Fig. 11. Plot of the feature matches that are used as loop closing constraints. (a) Before loop closing. (b) After loop closing.

References

1. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **20** (2003) 91–110
2. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24** (1981) 381–395
3. Nistér, D., Stewénius, H.: Scalable recognition with a vocabulary tree. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York City, New York. (2006) 2161–2168
4. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proc. 9th International Conference on Computer Vision, Nice, France. (2003) 1470–1477
5. Micusik, B., Pajdla, T.: Autocalibration and 3d reconstruction with non-central catadioptric cameras. In: CVPR 2004. (2004)
6. Micusik, B., Pajdla, T.: Structure from motion with wide circular field of view cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 1135–1149
7. Geyer, C., Daniilidis, K.: Structure and motion from uncalibrated catadioptric views. In: CVPR 2001. (2001)
8. Svoboda, T., Pajdla, T., Hlavac, V.: Motion estimation using central panoramic cameras. In: IEEE Int. Conf. on Intelligent Vehicles. (1998)
9. Chang, P., Hebert, M.: Omni-directional structure from motion. In: IEEE Workshop on Omnidirectional Vision. (2000)
10. Corke, P.I., Strelow, D., , Singh, S.: Omnidirectional visual odometry for a planetary rover. In: IROS. (2004)
11. Lhuillier, M.: Automatic structure and motion using a catadioptric camera. In: IEEE Workshop on Omnidirectional Vision. (2005)
12. Bosse, M., Rikoski, R., Leonard, J., Teller, S.: Vanishing points and 3d lines from omnidirectional video. In: ICIP02. (2002) III: 513–516
13. Bosse, M., Newman, P., Leonard, J., Teller, S.: Simultaneous localization and map building in large cyclic environments using the atlas framework. The International Journal of Robotics Research **23** (2004) 1113–1139

inria-00325324, version 1 - 28 Sep 2008

14. Bosse, M.: ATLAS: A Framework for Large Scale Automated Mapping and Localization. PhD thesis, Massachusetts Institute of Technology (2004)
15. Cummins, M., Newman, P.: Probabilistic appearance based navigation and loop closing. In: IEEE International Conference on Robotics and Automation (ICRA'07), Rome (2007)
16. Ho, K.L., Newman, P.: Detecting loop closure with scene sequences. International Journal of Computer Vision **74** (2007) 261–286
17. Scaramuzza, D., Siegwart, R.: Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. IEEE Transactions on Robotics, Special Issue on Visual SLAM **24** (2008)
18. Scaramuzza, D., Siegwart, R.: Monocular omnidirectional visual odometry for outdoor ground vehicles. In: 6th International Conference on Computer Vision Systems. (2008) 206–215
19. Triggs, B.: Autocalibration from planar scenes. In: ECCV98. (1998)
20. Labrosse, F.: The visual compass: performance and limitations of an appearance-based method. Journal of Field Robotics **23** (2006) 913–941
21. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle adjustment: A modern synthesis. In: Vision Algorithms Workshop: Theory and Practice. (1999) 298–372
22. Ni, K., Steedly, D., Dellaert, F.: Out-of-core bundle adjustment for large-scale 3d reconstruction. In: Proc. 11th International Conference on Computer Vision, Rio de Janeiro, Brazil. (2007) 1–8

inria-00325324, version 1 - 28 Sep 2008