Calibration Methodology for Distant Surveillance Cameras

Peter Gemeiner(✉), Branislav Micusik, and Roman Pflugfelder

AIT Austrian Institute of Technology GmbH, Vienna, Austria
peter.gemeiner@ait.ac.at

Abstract. We present a practical method for video surveillance networks to calibrate their cameras which have mostly non-overlapping field of views and might be tens of meters apart. The calibration or estimating the camera pose, focal length and radial distortion is an essential requirement in video surveillance systems for any further automated tasks like person tracking or flow monitoring. The proposed methodology casts the calibration as a localization problem of an image with respect to a 3D model which is built a priori with a moving camera. The method comprises state-of-the-art functioning blocks, the Structure from Motion (SfM) and minimal Perspective-n-Point (PnP) solvers, which were proved stable in 3D computer vision community and applies them in context of video surveillance. We demonstrate that the calibration method is effective in difficult repetitive, reflective and texture less large indoor environments like an airport.

Keywords: Video surveillance · Networked cameras · Calibration

1 Introduction

Surveillance camera networks of environments like airports, train stations, shopping malls etc, are backbones of security and monitoring systems to prevent hazardous situations and to guarantee smooth operation and people flow. The networks consist of high number of cameras, in magnitude thousands, in heavily non-overlapping setup, tens of meters apart cameras. The non-overlapping setup reflects the fact that typically all entrance and exit points are covered, but not entire areas which would yield enormous number of the cameras. Tasks like automatic visual tracking of people, person re-identification, people flow monitoring across such networked cameras are tasks which on one side would automate the operation processes significantly, but on the other, still represent substantial scientific challenges.

The aforementioned tasks can be solved better when the spatial mutual positions of all the cameras are available. Estimating position and orientation of

This research has been supported by funding from the Austrian Research Promotion Agency (FFG) project LOLOG nº 3579656 and PAMON nº 835916 and from EU FP7 under grant agreement nº 257906.

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part III, LNCS 8927, pp. 162–173, 2015.
DOI: 10.1007/978-3-319-16199-0_12

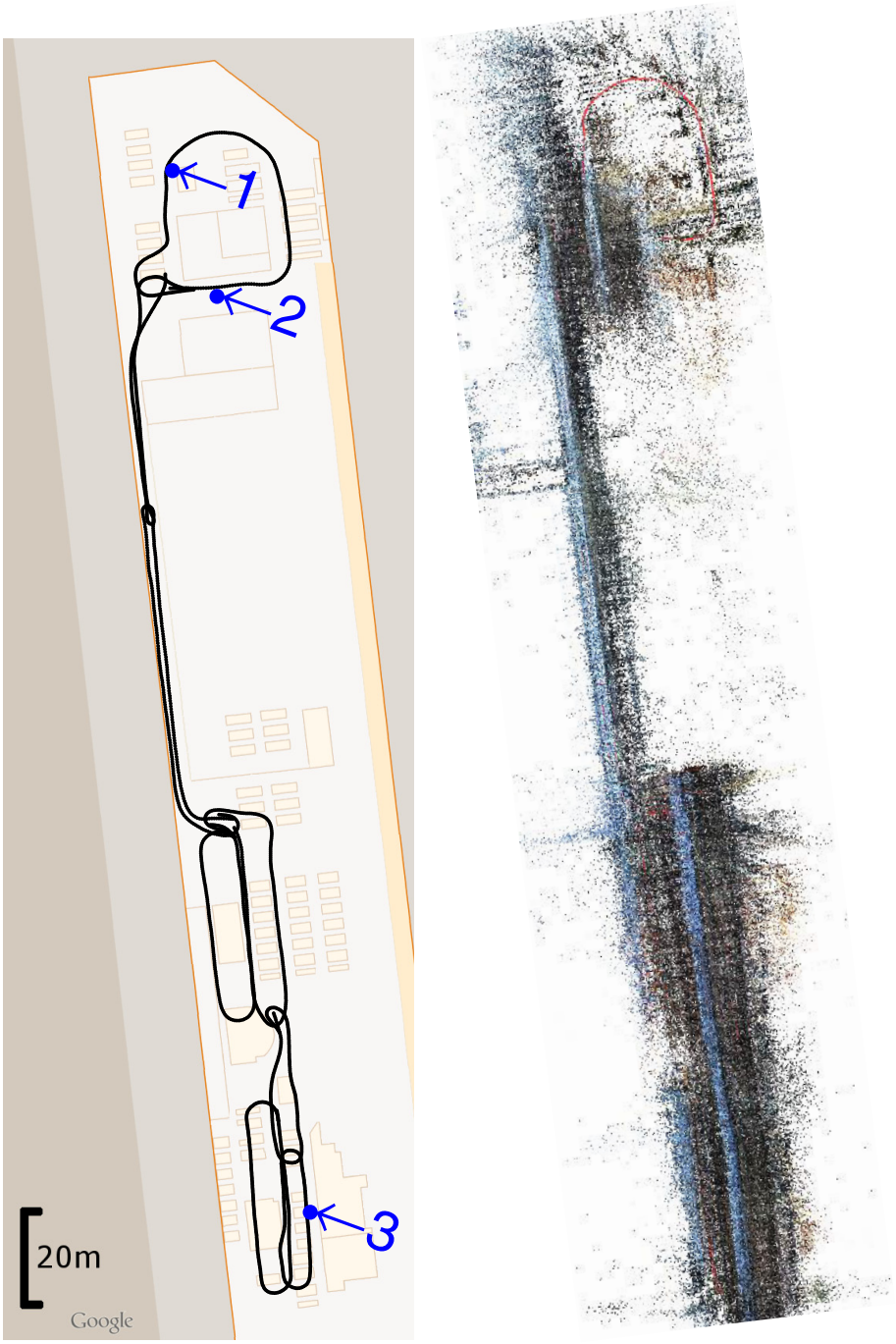

Fig. 1. Left side depicts three positions of calibrated surveillance cameras within the map of an airport terminal. Right side displays the reconstructed point cloud. This image is best viewed in color.

cameras which are tens meters apart, however, represent a significant challenge. No calibration target like a checkerboard can be used for estimating poses of cameras as the target cannot be seen in one time by more than one camera. To measure physically the distances of the cameras, and rotations e.g. by IMU, is cumbersome, very laborious and inaccurate. To use moving people for the calibration is rather theoretical than a practical concept. Therefore, there is no practically feasible methodology which would be demonstrated to work on large scale with acceptable effort of a user.

In this paper, we propose a practical method which was proved to be able to calibrate cameras which are 200m apart with accuracy of 1 meter in pose, as Fig. 1 shows. The main idea is to use as a calibration target the environment itself. It allows to bring all the cameras in one coordinate system and to estimate their internal parameters like the focal length and the radial distortion parameter in one go. The proposed methodology comprises state-of-the-art functioning blocks which were proved stable in 3D computer vision community in different context. In this paper, we merge those successful blocks, the SfM and minimal PnP solvers, and show in context of surveillance, that they both can be advantageously used for the calibration of distant networked surveillance cameras. Especially we demonstrate that calibration is effective in difficult repetitive, reflective and texture less large indoor environments.

2 Related Work

Calibration is theoretically sufficiently understood and the approach for estimation based on point and line correspondences is mature and well known [1][2]. Derived techniques for single surveillance cameras based on vanishing points [3][4], on foot-head homologies of visible persons [5][6] and on the motion of persons [7][8] exist. Nevertheless, surveillance companies still use simple calibration targets such as checkerboards in various sizes and sophisticated rigid rigs together with classic techniques based on point correspondences [9][10], as the reliability of the aforementioned techniques ist still not meeting the expectations.

This basic correspondence approach is generalisable to multiple camera views. A solution based on a laser pointer that is carried around by a person in the environment was shown by Svoboda et al. [11]. Funiak et al. [12] used a distinct marker carried by a person or robot. Deverajan et al. [13] presented a framework that takes the whole environment as calibration target. The environment's structure is assumed rigid and rich of matchable points. Despite all the success of these work, calibrating distant cameras, as the problem constitutes with a network of surveillance cameras, is surprisingly difficult and still a scientific challenge. The basic approach fails when cameras become more distant as either the point detection and matching breaks down or the camera views are disjoint.

To overcome these problems, a recent new approach [14][15][16] uses person tracking and exploits temporal continuity of the trajectory which can compensate the lack of correspondence. Unfortunately, it turned out that smoothness is a rather weak constraint compared to correspondence, hence the approach

becomes for larger distances between cameras infeasible (larger times persons are immeasurable), a situation that happens frequently in surveillance applications.

The proposed approach combines Deverajan's idea to use the whole environment as calibration target, but instead of solely using the images of distant cameras which clearly limits reliable matching, the approach enriches the set of images by a large collection taken with a portable camera. Similar work has been done in 3D reconstruction where techniques for camera localisation based on sparse 3D point clouds exist [17][18][19].

3 Approach

The methodology works as follows. First, we reconstruct the environment by visiting a place with an additional calibrated camera, acquiring a sequence in a free walk, loop like, trajectory. This step employs a standard SfM pipeline and reconstructs the scene as a sparse cloud of 3D points. Second, the surveillance cameras are stitched to the 3D model by the re-sectioning PnP (Perspective-n-Point) algorithm from 2D-3D correspondences.

3.1 Portable Camera

Before calibrating surveillance cameras in a new environment an image acquisition with a portable camera has to take place. This portable camera has known internal parameters and is equipped with 180° FOV optics as depicted in Fig. 2. This wide FOV proved to have two following advantages comparing to a standard camera. First, it can deliver longer feature tracks, which is important in poorly textured environments. Second, it provides well spatially distributed projective rays, which helps for estimating the epipolar geometry much more robustly.

Fig. 2. Portable camera equipped with wide FOV optics

3.2 Structure from Motion

After capturing the image sequence of the new environment with the portable camera the next step is the reconstruction of this environment. However, due to

the fact that the portable camera moved freely in space, without the usage of any additional sensors, its poses are unknown as well. Therefore the reconstruction of the unknown scene together with the camera poses leads to the classical SfM problem.

In recent years, several open-source SfM software packages appeared, e.g. Snavely's 'Bundler', which has roots in 'Photo tourism' [20] and the impressive 'Rome in a Day' project [21]. However, most of these SfM software packages include only perspective camera models, which is a dominant constraint for indoor spaces. The inclusion of an omni-directional camera model was the main reason why a custom SfM package has been developed within this work.

The custom SfM software used here contains the typical functional blocks:

- feature detection,
- establishing point correspondences,
- estimating pose between pairs of consecutive cameras,
- registration of all camera poses,
- triangulation of projective rays from point correspondences,
- loop-closing and
- non-linear optimization.

For feature detection several interest points together with their descriptors are selected (e.g. SIFT [22] and MSER [23]). The feature points are matched automatically in order to establish point correspondences, which are concatenated in tracks for more views. The matching process is assuming that the sequence of images is taken consecutively, which can help to reduce the amount of outliers in indoor spaces. However, the amount of outliers is still large and an epipolar geometry constraint is needed to validate them. For this validation the known five-point [24] algorithm is used and as a result the relative orientation and translation between a pair of images is obtained. With the help of these relative orientations and translations the projection matrices of all cameras are registered to the initial frame with linear method. In parallel to this, the reconstruction of the projective rays using the validated point correspondences is also linearly computed. To minimize the drift in scale, which typically appears in all odometry problems, the essential loop closing step is implemented as next step. With the help of this step additional point correspondences are found in images far apart from each other. In the last functional block, the linearly reconstructed points and camera poses are used for initialization for the non-linear optimization procedure. For this non-linear optimization the large framework called Bundle adjustment [25] is used, where as a cost function the reprojection error is selected.

3.3 Camera Calibration

Once the environment is reconstructed and represented as a sparse cloud of 3D points, the surveillance cameras are stitched to the 3D model with a re-sectioning algorithm. The re-sectioning, called also PnP (Perspective-n-Point) or absolute

positioning, stands for determining the absolute position and orientation of a camera from a set of 2D-to-3D point correspondences. It is one of the most important problems in computer vision with a solid theoretical background and a broad range of applications.

In context of surveillance, the following PnP algorithms are of interest. In most general case, where no information about the cameras is available, minimal non-linear P4Pfr [26,27] or non-minimal linear P5Pfr [28] solvers estimating camera position, orientation, focal length and radial distortion from four or five 2D-to-3D correspondences, respectively, can be utilized. If the internal calibration of the surveillance cameras is known, e.g. obtained by one of the checkerboard methods or through vanishing point estimation, then the P3P [29] is applicable. If e.g. a vertical vanishing point is detectable in the images of the surveillance cameras or a gravity vector of the camera is known, then vP3Pf [30] or gP3Pf [31], respectively, can be employed. In general, more information about the cameras is available, more accurate result can be achieved. In all our experiments we consider the most general P4Pfr case where no information about the cameras is available, showing thus the upper limit of the inaccuracy in calibration.

An important step in using the PnP algorithms is to establish the 2D-to-3D correspondences between the image of the surveillance camera and the 3D model. Typically, the sequence of a moving camera is acquired between one to two meters above the floor level. The surveillance cameras, however, are mounted couple of meters above the floor level to observe better the area from an elevated location and not to be reachable by people. This results in a very wide baseline setting of the cameras which reconstructed the 3D model and the surveillance cameras. Our experience showed that the difference in a viewpoint is mostly so large that none of the available image descriptors, e.g. [32], is invariant enough to handle it. To confirm that, we evaluated state-of-the-art approaches [17] [18] [19] for the large scale image based localization. They use effective strategies like visual words, vocabulary trees, prioritized search, virtual cameras, and perform reasonably well for localizing images which are spatially close to the images used for building the 3D model. However, in case of surveillance cameras the methods were ineffective in all our experiments. A fully automatic matching in this application domain therefore still remains a challenge.

Instead, we designed a simple GUI for establishing those 2D-to-3D correspondences manually. Experiments showed that only 6 to 10 correspondences are sufficient to get reasonable estimate and it is rather fast and not very laborious task. Despite that the matching is done manually, the whole concept shows its very high potential for calibrating purposes, indicating sufficient accuracy for following automated surveillance application like person tracking and people flow monitoring.

4 Experimental Results

The methodology presented here can be used for the calibration of surveillance cameras, which are mounted at larger distances from each other in different

environments, outdoor or indoor. In this paper, however, we show the most difficult case, the indoor scenario, as very difficult for SfM and PnP algorithms. Humans typically design indoor spaces so that they contain reflective surfaces, glass elements, narrow corridors, and almost no texture. In general, these properties cause problems to establishing point correspondences and bring most of the matching algorithms into failure. Despite that we demonstrate on real hard environments the possibility to cope with these problems and present a working method composed of our developed custom SfM software based on large FOV optics and of available PnP algorithms.

The custom SfM software has been tested in different indoor environments: two international airports and two office buildings. The image acquisition of these

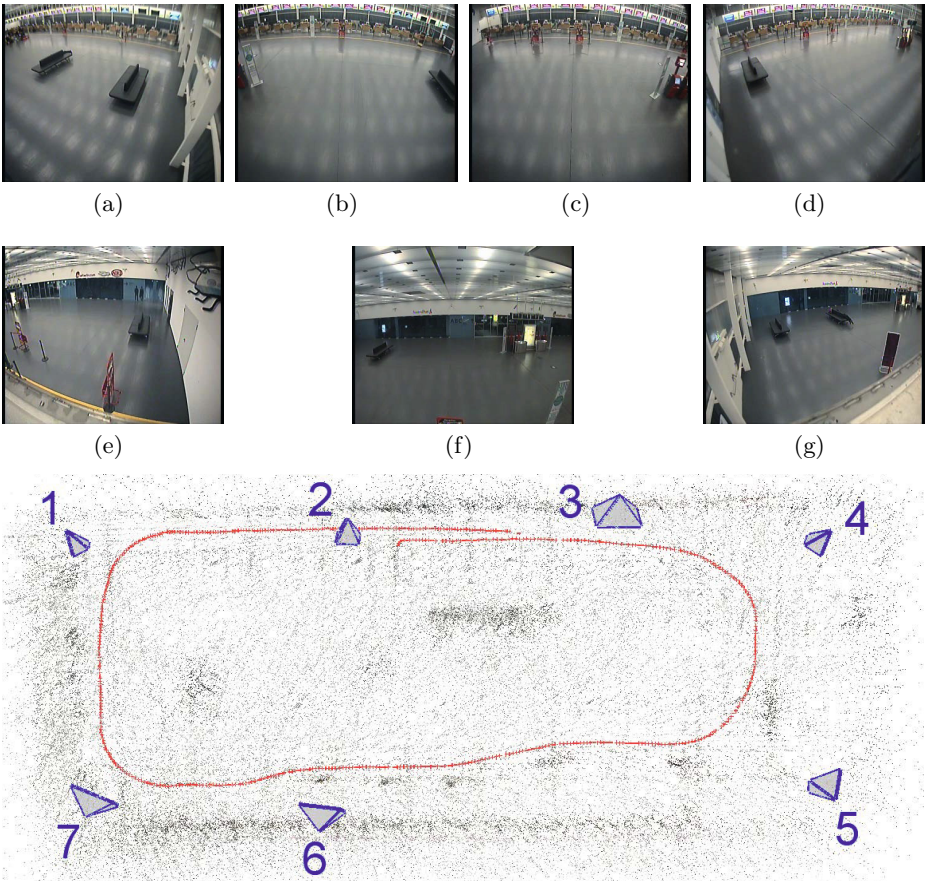

Fig. 3. The two top rows show seven views from the surveillance cameras at the Airport Vienna. The bottom image depicts the reconstructed environment as a sparse point cloud, the trajectory of the portable camera as a red line and the seven surveillance cameras calibrated with the presented methodology. This image is best viewed in color.

environments contains single loop captured in continues shutter mode with an partial overlap of the captured images and baseline of roughly 30cm. Ground Truth is not available for the presented sequences, therefore, as a way to quantitatively judge the result, we compare the estimated positions of the surveillance cameras to the wall or ceiling which are visible in the sparse 3D models. We roughly know where all the cameras are and how they align to each other. Moreover, for the airport sequence (see Fig. 1), to even better judge the results, we overlay the reconstructed trajectory and the position of the surveillance cameras with the Google Maps.

The first presented sequence is the Terminal from an airport. Fig. 1 shows the estimated position of the surveillance cameras and the sparse cloud of 3D points. This is the largest reconstruction in this paper, it contains eight partial reconstructions, stitched semi-automatically together. The covered area is about 40m x 200m. One can notice that the estimated trajectory of the moving camera is correctly estimated as it does not cross the walls and mainly follows the open space. We calibrate three surveillance cameras which are 200m apart and whose estimated poses roughly correspond to their true locations in the map. Due to security reasons, no screenshots from the surveillance cameras can be published.

The second sequence is from the international airport of Vienna, see Fig.3. This environment similarly to the previous sequence contains reflective surfaces,

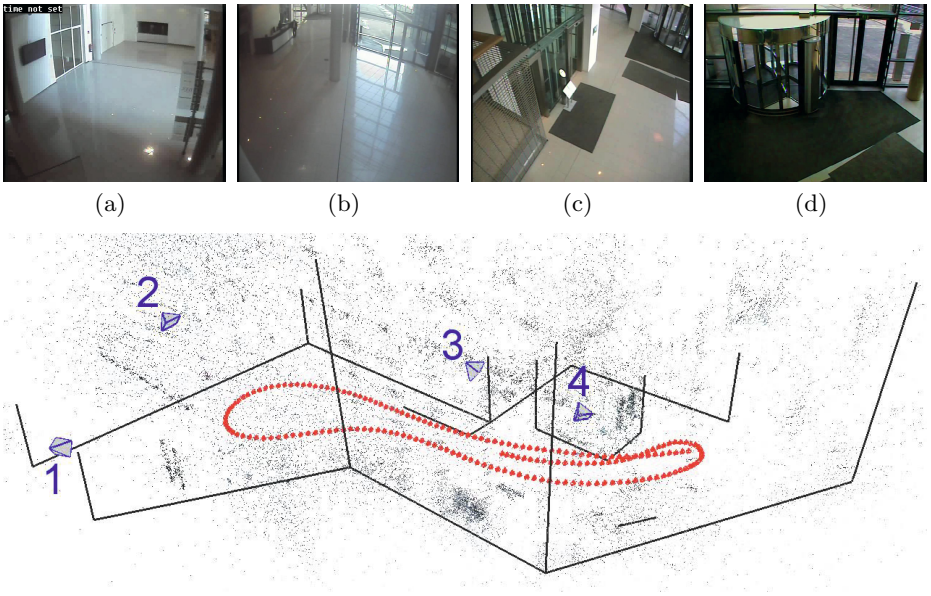

Fig. 4. The top row shows four views from the surveillance cameras in an office environment. The bottom image depicts the reconstructed environment as a sparse point cloud, the trajectory of the portable camera as a red line, and four surveillance cameras calibrated with the presented methodology. This image is best viewed in color.

very few textured objects, and almost one color floor, which make it very difficult for the SfM and calibration. We calibrated 7 cameras which are mounted in a way that the camera centers lie on two parallel lines aligned with two opposite walls. This was correctly reconstructed, as can be seen in Fig.3.

The next sequence is a hall of a modern office building depicted in Fig. 4 which is very similar to the large infrastructure buildings as e.g. airports or train stations. All the lines in the 3D model were additionally reconstructed through manual correspondences in order to improve visualization and better judge the estimated poses of the four surveillance cameras. All these cameras

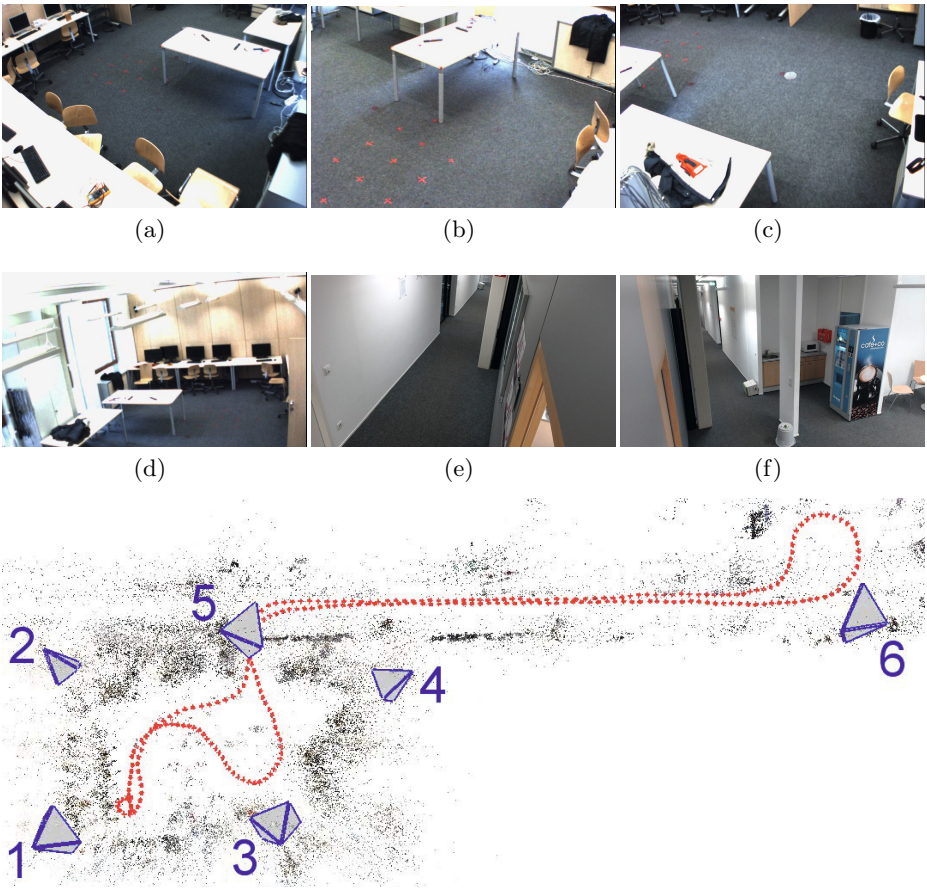

Fig. 5. The two top rows show six views from the surveillance cameras in an office / corridor environment. The bottom image depicts the reconstructed environment as a sparse point cloud, the trajectory of the portable camera as a red line, and six surveillance cameras calibrated with the presented methodology. This image is best viewed in color.

are aligned in reality but also in the estimation with the walls bounded by the reconstructed lines.

The last sequence is a narrow office corridor, shown in Fig. 5. Here, we had to visit one of the rooms as the surveillance cameras were mounted along the corridor and in the room as well. We had to enter and leave again the room through a door which is one of the most difficult situations in indoor SfM and SLAM community. The door causes separation of the room and the corridor spaces and it is hard to keep necessary amount of tracks for successful pose estimation. However, as the result shows, the reconstructed 3D structure is feasible and useful for further calibration of 6 cameras. The estimated positions of the six cameras visually align with their true locations.

To summarize, for all the sequences from various environments, our comparisons indicate accuracy in the pose estimation of the surveillance cameras to be under 1m. This is far sufficient for further tasks like multi cameras person tracking, people flow monitoring across networks of distant cameras.

5 Conclusions

We presented a practical methodology for calibrating very distant surveillance cameras, apart tens of meters, in difficult indoor environments, poorly textured, full of repetitiveness, and reflections. We show that building on well researched parts in geometry community of computer vision area, namely Structure from Motion and Perspective-and-Point algorithms, a sufficiently accurate and practically interesting method can be brought into the video surveillance field. The method does not require a laborious placement of any calibration target, and when managing fully automatic matching, a large number of cameras could be conveniently calibrated. To the best of our knowledge, this is the first work which demonstrates a calibration method on large airport scenario with cameras hundreds meters apart.

Acknowledgements. Authors would like to thank Gustavo Fernandez and Georg Nebehay for the help with the image acquisition. Authors are also gratefull to the smart camera lab at University of Klagenfurt for their help with the experiments.

References

1. Faugeras, O., Luong, Q.T., Papadopoulou, T.: The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications. MIT Press (2001)
2. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004)
3. Wildenauer, H., Hanbury, A.: Robust camera self-calibration from monocular images of manhattan worlds. In: Proc. CVPR (2012)
4. Wildenauer, H., Micusik, B.: Closed form solution for radial distortion estimation from a single vanishing point. In: Proc. BMVC (2013)

5. Krahnstoever, N., Mendonca, P.: Bayesian autocalibration for surveillance. In: Proc. ICCV (2005)
6. Micusik, B., Pajdla, T.: Simultaneous surveillance camera calibration and foothead homology estimation from human detections. In: Proc. CVPR (2010)
7. Rahimi, A., Dunagan, B., Darrell, T.: Tracking people with a sparse network of bearing sensors. In: Pajdla, T., Matas, J.G. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 507–518. Springer, Heidelberg (2004)
8. Krahnstoever, N., Mendona, P.R.S.: Autocalibration from tracks of walking people. In: Proc. BMVC (2006)
9. Heikkila, J., Silven, O.: A four-step camera calibration procedure with implicit image correction. In: Proc. CVPR (1997)
10. Zhang, Z., Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**, 1330–1334 (1998)
11. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multicamera self-calibration for virtual environments. Presence: Teleoperators and Virtual Environments **14**(4), 407–422 (2005)
12. Funiak, S., Guestrin, C., Paskin, M., Sukthankar, R.: Distributed localization of networked cameras. In: Proc. IPSN (2006)
13. Devarajan, D., Cheng, Z., Radke, R.: Calibrating distributed camera networks. Proceedings of the IEEE **96**(10), 1625–1639 (2008)
14. Rahimi, A., Dunagan, B., Darrell, T.: Simultaneous calibration and tracking with a network of non-overlapping sensors. In: Proc. CVPR (2004)
15. Rudoy, M., Rohrs, C.: Enhanced simultaneous camera calibration and path estimation. In: Proc. ACSSC (2006)
16. Picus, C., Pflugfelder, R., Micusik, B.: Auto-calibration of non-overlapping multicamera CCTV systems. In: Shan, C., Porikli, F., Xiang, T., Gong, S. (eds.) Video Analytics for Business Intelligence. SCI, vol. 409, pp. 43–67. Springer, Heidelberg (2012)
17. Irschara, A., Zach, C., Frahm, J.M., Bischof, H.: From structure-from-motion point clouds to fast location recognition. In: Proc. CVPR (2009)
18. Sattler, T., Leibe, B., Kobbelt, L.: Fast image-based localization using direct 2D-to-3D matching. In: Proc. ICCV (2011)
19. Sattler, T., Leibe, B., Kobbelt, L.: Improving image-based localization by active correspondence search. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part I. LNCS, vol. 7572, pp. 752–765. Springer, Heidelberg (2012)
20. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. ACM Transactions on Graphics (SIGGRAPH Proceedings) **25**(3), 835–846 (2006)
21. Agarwal, S., Snavely, N., Simon, I., Seitz, S., Szeliski, R.: Building rome in a day. In: International Conference on Computer Vision (2009)
22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**, 91–110 (2004)
23. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference, pp. 384–393 (2002)
24. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Pattern Analysis and Machine Intelligence **26**(6), 756–770 (2004)
25. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) ICCV-WS 1999. LNCS, vol. 1883, pp. 298–372. Springer, Heidelberg (2000)

26. Josephson, K., Byröd, M.: Pose estimation with radial distortion and unknown focal length. In: Proc. CVPR (2009)
27. Bujnak, M., Kukelova, Z., Pajdla, T.: New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part I. LNCS, vol. 6492, pp. 11–24. Springer, Heidelberg (2011)
28. Kukelova, Z., Bujnak, M., Pajdla, T.: Real-time solution to the absolute pose problem with unknown radial distortion and focal length. In: Proc. ICCV (2013)
29. Kneip, L., Scaramuzza, D., Siegwart, R.: A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In: Proc. CVPR (2011)
30. Micusik, B., Wildenauer, H.: Minimal solution for uncalibrated absolute pose problem with a known vanishing point. In: Proc. 3DV (2013)
31. Kukelova, Z., Bujnak, M., Pajdla, T.: Closed-form solutions to minimal absolute pose problems with known vertical direction. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part II. LNCS, vol. 6493, pp. 216–229. Springer, Heidelberg (2011)
32. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2), 91–110 (2004)