Automatic Reconstruction of Wide-Area Fiducial Marker Models

Manfred Klopschitz* Dieter Schmalstieg

Graz University of Technology

ABSTRACT

We present an approach towards automatic reconstruction of large assemblies of fiducial markers scattered throughout a wide indoor area, using a computer vision based reconstruction approach. The data is acquired from a video stream captured with a monoscopic camera. The system is capable of creating markers models that are significantly larger in physical area and number of markers than with previous approaches.

1 INTRODUCTION

In 2000, Klinker et al. [15] proposed the AR-ready building, describing a vision of globally outfitting edifices with infrastructure for AR tracking. Indeed, most tracking systems that are successfully deployed in practice rely on some kind of environmental infrastructure. Passive fiducial markers are popular because they are easily available and provide reliable inside-out tracking with the limited hardware resources of a wearable single-camera computer. Overall, a marker-based tracking system is very inexpensive, and numerous markers can be deployed throughout the tracking area without having to consider tethering or power supply issues. This has prompted the development of special hardware solutions such as the Intersense IS1200 [19].

While deploying markers is straight forward, obtaining a global calibrated model of the markers is much more difficult. Unfortunately, building a global marker model cannot be avoided since its availability is essential for providing self-localization and tracking throughout the entire work area. In fact, in cases where obtaining a detailed CAD model of the environment for model based tracking is infeasible, a marker model may be the only feasible approach for stable tracking in a large environment. The problem of obtaining a marker model is acerbated if the marker assembly is large and dispersed throughout a wide non-homogeneous area, so that it is impossible to observe all markers simultaneously.

We would like to build such a global marker model automatically and with minimal human assistance. A video stream of the environment is captured with a calibrated video camera. Images from this video stream are then used to extract salient images containing both artificial and natural features. By using both types of features, sparse marker configurations can also be handled successfully – a prerequisite for deploying the system in practical situations where an excessive amount of markers may not always be possible.

In this paper, we present a method for ergonomic creation of large scale marker models. A calibrated video camera was used to obtain video sequences from a non-homogeneous indoor area decorated with markers. Salient images from the sequence were selected for camera path reconstruction, followed by a global bundle adjustment of the position of markers identified in the video. The result was used for inside-out tracking using a mobile handheld computer with a single camera. Our preliminary results indicate that the work flow described in this paper is suitable for obtaining larger global marker models than previously possible. We also discuss current limitations in the reconstruction pipeline and how they can be addressed with future work.

*e-mail: klopschitz@icg.tugraz.at

[0]This work was sponsored by the Austrian Science Fund under contracts Y193 and W1209-N15.

2 RELATED WORK

2.1 Augmented Reality

Marker-based based tracking techniques such as ARToolKit [14], ARTag [7] and ARToolKitPlus [30] have turned out to be extremely popular in Augmented Reality. Despite significant advances in natural feature tracking [10], marker-based tracking is more easily available, more computationally efficient, more reliable with respect to initialization and does not require a CAD model of the environment. For simple, small scale cases such as a printed poster board with multiple markers, the marker model is usually given by design.

However, Baratoff et al. [3] argue that there is a class of applications that require to instantly deploy a larger collection of markers in a medium-size environment and create a calibrated multi-marker model with minimal human assistance. They suggest a technique for rapidly computing a global marker model from multiple images of the markers, based on global bundle adjustment of the feature points (corners) of the markers. An advantage of this method is that the technique can be performed with the same hardware (camera, computer) used later for the online pose tracking, and computation runs in near real-time – they report 1 sec for 12 markers and 40 images.

A disadvantage of the method is that a high amount of redundancy is required, and the reconstruction of the marker model is only guaranteed to be stable if several images show all or nearly all of the markers. Larger marker assemblies have in the past been calibrated by hand, for example using a tachyometer [24]. Other approaches for geometric input directly in AR such as [23] and [2] seem not suitable for the efficient creation of marker models. Clearly, manual model creation is too labour intensive in practice and becomes infeasible after a certain point.

Our approach is an improvement over [3] in that key frame images are extracted automatically from a video stream, and that large areas covered by sparse markers can be handled by considering not only artificial markers, but also naturally occluding features for the reconstruction.

2.2 Computer Vision

Determining positions of markers in a 3D coordinate frame from 2D images or a video stream is a classical Structure from Motion (SfM) problem. There exists a large body of work in computer vision and photogrammetry on this topic. A good overview of basic techniques can be found in [13]. One common classification approach for SfM algorithms is to differentiate between offline batch strategies and sequential approaches.

Examples of batch systems are [9] and [21], where images are obtained in sequential order but processed in a hierarchical way. These approaches try to distribute the accumulated error optimally over the whole sequence. The sequential order of input images reduces the feature matching problem to a tracking problem. An approach for unordered sets of input images is presented in [26] where

images are matched using SIFT (Scale Invariant Feature Transform) [16] features. Images are added to the reconstruction sequentially, the ordering strategy is based on criteria to maximize reconstruction stability. Another strategy for reconstructing unordered sets of images can be found in [27]. This method constructs a minimal spanning tree of a camera adjacency graph with edges approximating geometric proximity.

Vision based SLAM (Simultaneous Localization and Mapping) methods compute the current camera position and orientation in real-time. Incremental map building and continuous localization increase the robustness of SfM as demonstrated in [5]. The area that can be covered is limited by the number of landmarks that can be recognized efficiently and the optimization techniques possible in real-time to increase accuracy. An alternative method to obtain an odometry of a moving camera is presented in [22]. Long camera trajectories can be obtained with this method, but a map is not constructed. This means that the algorithm can not recover from a camera movement in the sequence that is not suitable for reconstruction.

Bundle adjustment [28] is used in many systems as a global optimization to obtain an accurate reconstruction after an initial estimate of the scene. The main limitation of bundle adjustment is the computing time for long image sequences. A method that applies bundle adjustment in an incremental way suitable for long reconstruction sequences is presented in [18]. Here the weak geometric redundancy over a long image sequence typically found in odometry applications is exploited to combine a visual odometry approach with online bundle adjustment.

3 Marker Reconstruction Framework

To calibrate the position of fiducial markers in 3D space a video camera is used as a measurement device. The camera moves around in space and records a video stream containing the markers. When the SfM problem for this sequence is solved, markers are detected using ARToolKitPlus. SfM gives the camera position and orientation for each reconstructed image from the video. This means a marker position can be obtained simply by triangulation with all views where a marker is visible. A rigid transformation can then be used to place the marker positions in an absolute world coordinate frame. The overall SfM strategy is similar to the system described in [18].

Only calibrated cameras with known intrinsic parameters are used to obtain a reconstruction. Although an uncalibrated setup is more flexible, the main reason for a calibrated one is to increase the robustness, especially for planar structure dominated scenes. To obtain image point correspondences natural features are tracked in the video stream. Relying on fiducial markers in the SfM stage would mean that markers have to be always visible. In addition, accurate SfM results depend on a high number of well distributed feature points. These requirements make a simpler marker based SfM implementation not suitable for most practical situations.

Not all images of a video stream can be used for reconstruction because the parallax between consecutive frames is not sufficient for triangulation. Therefore an image from the video is inserted into the reconstruction only when this requirement is fulfilled.

3.1 Feature Tracking

For feature matching a sparse optical flow KLT (Kanade-Lucas-Tomasi) feature tracker is used. The implementation is provided by *OpenCV* [4]. Natural features are extracted using a classical corner detector [25]. These feature points are tracked until a threshold of lost points is reached. Then new corners are extracted and merged with existing ones. The number of lost tracks is a good indicator of motion parallax (ignoring pure rotation). It is used as a simple but effective heuristic to determine video frames suitable for reconstruction. Here an important detail is to monitor feature track statistics over three views because SfM algorithms need correspondence information over at least this number of views to integrate them into one reconstruction. Figure 1 shows an example of the natural feature tracking.

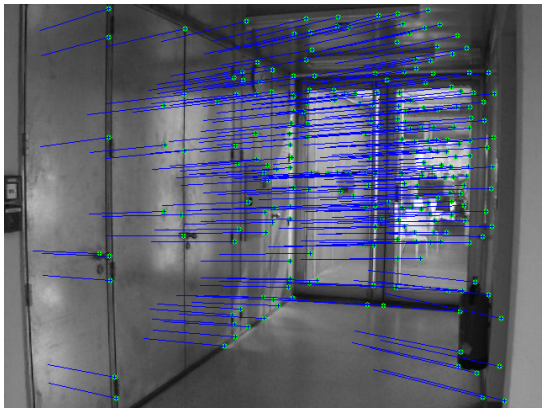

Figure 1: Natural feature tracking. Green points indicate corner locations and the blue lines the positions in the last frame used for reconstruction.

3.2 Reconstruction

3.2.1 Initialization

A 3D coordinate frame is initialized with the first three views of the sequence. The relative orientation is computed between the first and third image with the five-point algorithm, as described in [20]. The center view is inserted using a classical nonlinear three-point absolute pose algorithm [11], the implementation is based on a resultant method [1]. This scheme has the advantage that it works in planar and general scene cases without model selection. The minimal solution algorithms are used in RANSAC (Random Sample Consensus) loops [8] to determine an inlier set of point correspondences. Bundle adjustment is then used to obtain a least squares estimate.

3.2.2 Adding a View

After an initial reconstruction is available new cameras are inserted with a RANSAC three-point absolute pose algorithm. Feature points in this stage have to be visible in at least two previously reconstructed cameras. The points are triangulated and provide the 3D structure to compute the new absolute orientation. For the triangulation two cameras with a good baseline are selected. In this stage an optimal (two view) triangulation algorithm [12] is used. To obtain a least squares solution bundle adjustment is carried out over the last n cameras, with $n = 10$ and the oldest $n_c = 7$ cameras held constant typically. This is justified by [18] and is consistent with our own experiments.

3.2.3 Bundle Adjustment

Bundle adjustment is an essential step to provide a least squares estimate for solutions of minimal algorithms and increases significantly the stability of this sequential reconstruction approach. In addition, it allows to add weaker correspondences that are only visible in two cameras to the adjustment computations. Given the reconstruction problem

$$\mathbf{x}_j^i = P^i \mathbf{X}_j, \tag{1}$$

where the 2D point measurements $\mathbf{x}_j^i$ are the observations of unknown 3D points $\mathbf{X}_j$ observed in the unknown cameras P^i, bundle adjustment is defined as the minimum of this cost function:

$$\mathscr{C}(P^i, \mathbf{X}_j) = \sum_{ij} d(P^i \mathbf{X}_j, \mathbf{x}_j^i)^2. \tag{2}$$

This non-linear optimization problem is typically solved with the Levenberg-Marquardt algorithm. In our Euclidean case the number of unknowns per camera is 7 (4 for quaternion rotation parametrization and 3 for position) and 3 for each 3D point. To solve this optimization problem efficiently, the sparse structure of the problem (in short, not all points are visible in all cameras) can be exploited, and the partial derivatives of the cost function can be precomputed symbolically. The software package by [17] provides a basic framework for this optimized approach.

3.2.4 Marker Reconstruction

After the SfM problem is solved, each marker detected in the video stream is triangulated. A marker is triangulated using a linear n-view method with all available cameras. Then the structure is optimized by minimizing the re-projection error in all images.

3.2.5 Work Flow

The steps necessary to obtain a fiducial marker tracking model with our framework can be summarized as follows:
Marker placement: Place fiducial markers in the area of interest. There are no special constraints.
Video recording: A camera is used to record a video stream covering all markers. To obtain optimal results the camera trajectory should be chosen so that the parallax (with respect to the fiducial markers) between views is maximized.
Feature tracking and reconstruction: The video stream is processed with our framework. Natural features are tracked in the video to select views for reconstruction. Marker positions are detected and stored for these frames. The SfM problem is solved and 3D marker positions are triangulated using all available views.
AR model alignment: The result of the previous step is a 3D model of the marker configuration in an unknown coordinate frame. To align the tracking model with an AR model a rigid body transformation [6] is computed with at least three 3D point correspondences. The tracking model is then transformed into the AR target coordinate frame and a marker configuration file for ARToolKitPlus is created.

4 Applications and Results

We present two automatic marker reconstruction examples. A fully working AR application is used to demonstrate the applicability of the whole work flow and a larger scenario shows the scalability and flexibility gained by natural feature tracking over previous methods.

4.1 Vidente Indoor AR Application

Vidente [29] is a mobile augmented reality application for the 3D-visualization of subsurface features like power lines and water pipes. For indoor demos an architectural model and marker tracking is utilized. Figure 2 shows the indoor application using automatically reconstructed markers and illustrates the SfM intermediate step.

A manually created, highly accurate calibration is available for this model. Using our Vidente indoor demo with the manual and automatic marker calibration data showed no visually detectable difference in tracking quality. The physical size of the board is $120 \times 90cm$. For all 17 markers the mean center position error is $0.8cm$ with a standard deviation of $\sigma = 0.47cm$.

4.2 Large Scale Scenario

We demonstrate the handling of physically large scenarios and sparse, irregular marker placement where only natural feature information is available in parts of the sequence. The longest wall is $9m$ wide. Fiducial landmarks are placed with varying distances. Figure 3 shows the scene and its marker reconstruction.

The number of present fiducials is in general not a limiting factor. The main challenges are to handle build up of drift in the SfM step and the creation of a good camera trajectory where the parallax is maximized and the markers are detectable in the images. This

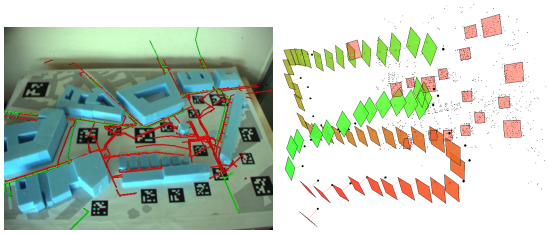

Figure 2: Vidente indoor demo and automatic marker reconstruction. Camera order is indicated by color, from red to green. Marker positions are highlighted by red rectangles.

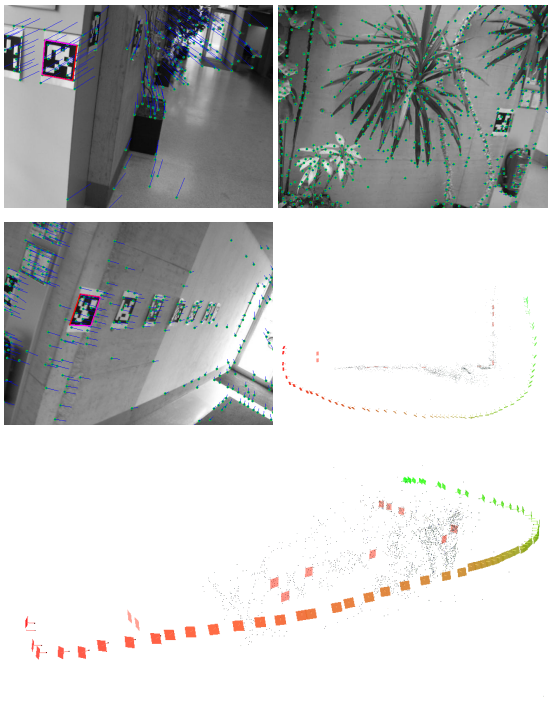

Figure 3: Large scale scenario images and reconstruction. The length of the longest wall segment is $9m$.

buildup of drift and its influence on the absolute reconstruction error is demonstrated in Figure 4. A $15m$ long trajectory with markers placed in a regular one meter interval along a straight line is used as ground truth. In this drift evaluation sequence the mean marker center position error is $0.17m$ with a standard deviation of $\sigma = 0.14m$. The largest error is at the end of the sequence with 0.44m deviation from the ground truth. As expected for odometry, the overall error tendency is super-linear with respect to the sequence length.

There is no restriction for marker placement as long as the sequence observing the scene is appropriate for the computer vision tasks. The run time performance of the system is near real-time and does not degrade with the length of the sequence. This means that the limiting factors are odometry drift and the availability of natural (or fiducial) features. In practice the most difficult part is to provide a SfM suitable motion.

5 Limitations

The used SfM approach depends on the continuity of the video sequence. If a marker is not detected during the tracking stage and missing in the reconstruction, it can not be merged easily with another video stream. This feature could be added by combining different marker reconstructions into one coordinate frame.

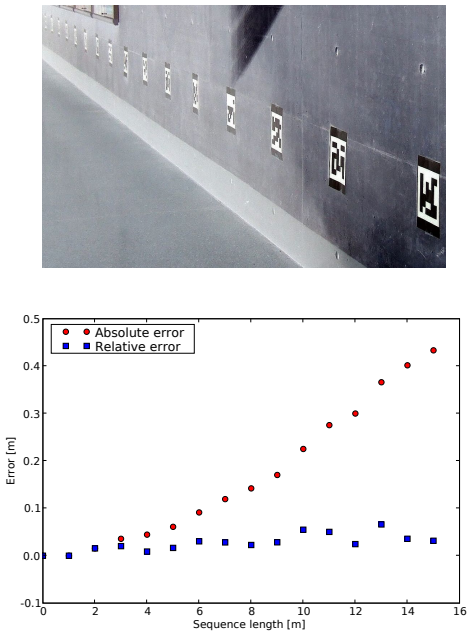

Figure 4: SfM drift influence on marker reconstruction. Markers are placed in a regular one meter interval along a $15m$ long straight line. Absolute error: Accumulated marker position error. Relative error: Deviation from the regular one meter marker placement.

A more difficult problem is how camera rotation is handled. Pure rotation without parallax is a typical limitation of monocular reconstruction approaches. For this problem, the most promising solution at the moment are SLAM methods. These methods can use previously reconstructed landmarks to recover from failure.

Like the tracking stage, the geometry computation pipeline considers only sequential camera dependencies. The accuracy can be enhanced if loop closing possibilities are recognized. The accuracy evaluation itself is a topic that can be improved. There are two obvious evaluation strategies: 3D reconstruction error evaluation with ground truth and judging the pose tracking performance in an AR application.

6 Conclusion

This paper presents an approach towards automatic and flexible large scale fiducial marker calibration. A calibrated camera is used as a measuring device to create models for visual tracking automatically. Computer vision algorithms provide the basic foundation for this. We tested the concept and work flow with a real AR application and create a marker model for a large scale setup.

The used SfM approach can be applied directly to AR applications. Visual pose tracking in low marker density areas could be supplemented with a real time SfM method. In such a setting a hybrid visual pose tracker could be used.

Traditional marker tracking provides robust and effective tracking information while SfM techniques can be used to build tracking models automatically or reduce the necessary fiducial landmark density.

References

[1] M. Ameller, M. Quan and B. Triggs. Camera Pose Revisited New Linear Algorithms. In *Rapport Interne - Equipe MOVI, 2000.*

[2] Y. Baillot and D. Brown and S. Julier. Authoring of Physical Models Using Mobile Computers. In *Proc. ISWC 2001, pages 39–46.*

[3] G. Baratoff, A. Neubeck and H. Regenbrecht. Interactive Multi-Marker Calibration for Augmented Reality Applications. In *Proc. ISMAR 2002, pages 107–116.*

[4] G. Bradski. Programmer's tool chest. The OpenCV library. In *Dr. Dobbs Journal, November 2000.*

[5] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc. ICCV, 2003, pages 1403–1410.*

[6] D. Eggert and A. Lorusso and R. Fisher. Estimating 3-D rigid body transformations: a comparison of four major algorithms. In *Mach. Vision Appl. J., volume 9, number 5–6, pages 272–290, 1997.*

[7] M. Fiala. ARTag, a fiducial marker system using digital techniques. In *Proc. CVPR, pages 590–596, 2005.*

[8] M. Fischler and R. Bolles:. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In *Comm. of the ACM, volume 24, pages 381–395, 1981.*

[9] A. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. ECCV, pages 311-326, 1998.*

[10] Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar and N. Navab. Marker-less Tracking for AR: A Learning-Based Approach In *Proc. ISMAR, pages 295–314, 2002*

[11] R. M. Haralick, C. Lee, K. Ottenberg and M. Nolle. Review and analysis of solutions of the three point perspective pose estimation problem. In *Int. J. Comput. Vision, volume 13, issue 3, pages 331–356, 1994.*

[12] R. Hartley and P. Sturm. Triangulation. In *Int. J. of Comput. Vis. and Image Underst., volume 68, issue 2, pages 146–157, 1997.*

[13] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision." Cambridge University Press, 2004, Second Edition

[14] H. Kato and M. Billinghurst. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. In *Proc. IWAR, pages 85–94, 1999.*

[15] G. Klinker, T. Reicher and B. Bruegge. Distributed User Tracking Concepts for Augmented Reality Applications. In *Proc. ISAR, pages 37-44, 2000.*

[16] D. Lowe. Distinctive image features from scale-invariant keypoints. In *Int. J. of Computer Vision, volume 60, issue 2, pages 91–110, 2004.*

[17] M. Lourakis and A. Argyros. The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package. In *TR 340, Inst. of Comp. Sci. - FORTH Heraklion, Crete, Greece, Aug. 2004. Available from* `http://www.ics.forth.gr/~lourakis/sba`

[18] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, Patrick Sayd. 3D Reconstruction of Complex Structures with Bundle Adjustment: an Incremental Approach. In *Proc. ICRA, pages 3055–3061, 2006.*

[19] L. Naimark and E. Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *Proc. ISMAR, pages 27- 36, 2002.*

[20] D. Nistér. An efficient solution to the five-point relative pose problem. In *Trans. PAMI, volume 26, issue 6, pages 756–770, 2004.*

[21] D. Nistér. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. *Proc. ECCV, pages 649–663, 2000.*

[22] D. Nistér., O. Naroditsky and J. Bergen. Visual odometry. In *Proc. CVPR, pages 652–659, 2004.*

[23] W. Piekarski and B. H. Thomas. Augmented reality working planes: A foundation for action and construction at a distance. In *Proc. ISMAR, pages 162–171, 2004.*

[24] G. Schall, F. Fraundorfer, J. Newman and D. Schmalstieg. Construction and Maintenance of Augmented Reality Environments Using a Mixture of Autonomous and Manual Surveying Techniques. In *Proc. 7th Conference on Optical 3-D Measurement Techniques, 2005.*

[25] J. Shi and C. Tomasi. Good Features to Track. In *Proc. CVPR, pages 593–600, 1994.*

[26] N. Snavely, S. Seitz and R. Szeliski. Photo tourism: exploring photo collections in 3D. In *Proc. SIGGRAPH, pages 835–846, 2006.*

[27] K. Steele and P. Egbert. Minimum Spanning Tree Pose Estimation. In *Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), pages 440–447, 2006.*

[28] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon. Bundle Adjustment – A Modern Synthesis. In *Lecture Notes in Computer Science, volume 1883, pages 298–372, 2000.*

[29] Vidente project (visualization of subsurface features). URL `http://studierstube.icg.tu-graz.ac.at/vidente/`

[30] D. Wagner and D. Schmalstieg. ARToolKitPlus for Pose Tracking on Mobile Devices. In *Proc. CVWW, pages 139–146, 2007.*