A Novel Approach to On-Site Camera Calibration and Tracking for MR Pre-visualization Procedure

Wataru Toishita, Yutaka Momoda, Ryuhei Tenmoku, Fumihisa Shibata,
Hideyuki Tamura, Takafumi Taketomi, Tomokazu Sato, and Naokazu Yokoya

Ritsumeikan University, Nara Institute of Science and Technology, Japan
toishita@rm.is.ritsumei.ac.jp

Abstract. This paper presents camera calibration and tracking method for mixed reality based pre-visualization system for filmmaking. The proposed calibration method collects environmental information required for tracking efficiently since the rough camera path and target environment are known before actual shooting. Previous camera tracking methods using natural feature are suitable for outdoor environment. However, it takes large human cost to construct the database. Our proposed method reduces the cost of calibration process by using fiducial markers. Fiducial markers are used as reference points and feature landmark database is constructed automatically. In shooting phase, moreover, the speed and robustness of tracking are improved by using SIFT descriptor.

Keywords: Mixed Reality, Pre-visualization, Tracking, Natural Feature.

1 Introduction

Geometric registration is the most significant issue for mixed reality (MR) which can merge real and virtual worlds. This issue results in the real-time tracking problem of feature points in the real world. It's not an exaggeration to say that a large part of MR research focuses solving this problem. In early days, the most common approach is arranging fiducial markers in the real world in order to realize robust detection and tracking of feature points. On the other hands, a lot of markerless methods using natural feature points are popular in recent years [1]. In tracking techniques, some powerful methods are proposed recently [2,3]. However, there is no definitive method in the initial camera calibration, therefore, there is no all-round total geometric registration method.

We address MR-based pre-visualization in filmmaking (MR-PreViz, Fig.1) as an application of MR technology [4,5]. The purpose of MR-PreViz is pre-designing camerawork by superimposing the computer generated humans and creatures onto the real movie sets. To say in filmmaking terms, it corresponds to the real-time on-set 3D matchmove in the pre-production process of filmmaking. We develop software tools supporting this process. Here, the camera calibration and tracking play an important role.

We develop an on-site camera calibration and tracking method suitable for this purpose. Our proposed method cleverly innovates the initial camera calibration using fiducial markers and the markerless tracking technique. The basis of this method is

R. Shumaker (Ed.): Virtual and Mixed Reality, LNCS 5622, pp. 492–502, 2009.

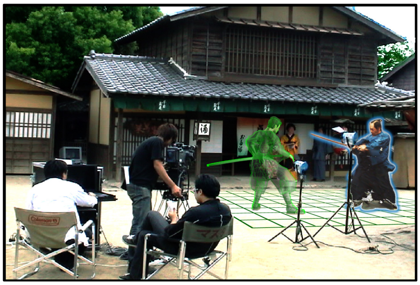


Fig. 1. Conceptual image of MR-PreViz

the vision-based 6DOF tracking method using the landmark database constructed in advance by detecting natural feature points in the real scene [6]. We improved this method based on that MR-PreViz has the setup and the shooting phases. Our proposed method is a sustainable approach for the high-speed and complicated camera work of the actual shooting.

This paper is constructed as follows. Section 2 describes the outline of the proposed method. Sections 3 and 4 show the initial camera calibration and tracking techniques of the proposed method, respectively. Finally, Section 5 summarizes this paper.

2 New Tracking Method Suitable for MR-PreViz Shooting

2.1 Geometric Registration in MR-PreViz

Real-time estimation of camera position and pose is needed in MR-PreViz. Here, the following conditions are required.

- System performs in the outdoor and indoor environments.
- It doesn't take so much time and human costs to collect any environmental information.
- Tracking must be realized without any fiducial markers.

Considering these conditions, we decided to adopt Taketomi's method [6] to MR-PreViz. In the setup phase, we construct the landmark database of the MR-PreViz shooting site (e.g. the location sites, open sets, and indoor sets) before shooting MR-PreViz movies. In MR-PreViz shooting phase, the camera position and pose are estimated (calibrated and tracked) in real-time using the constructed landmark database. Taketomi's method [6] is the most appropriate method to fill the above conditions. This approach can estimate extrinsic camera parameters from the captured images using correspondences between landmarks of the database and natural feature points in the input image. The landmark database stores 3D positions of natural feature points and image templates around them.

However, there still remain the following improvements in simply adopting the previous method [6] to MR-PreViz procedures.

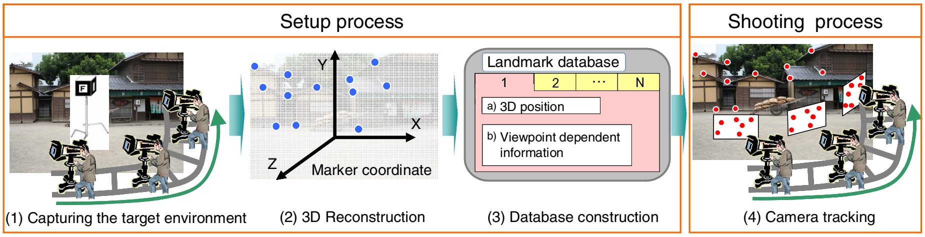


Fig. 2. Procedure of proposed method

- To reduce the human cost and computational time in constructing the landmark database
- To reduce the computational time of the initial camera calibration and tracking

In order to realize them, we propose the procedures that are described in the following section.

2.2 Procedures of Geometric Registration

The geometric registration procedures of the MR-PreViz system are shown in Fig.2. In the setup phase, as the first step, image sequences capturing the target environment in almost the same camera path as the MR-PreViz movie. In this step, the fiducial markers are arranged in the target environment to realize the robust and fast estimation of the camera path. Arranging fiducial markers makes it possible to reduce major part of the human cost. As the second step, 3D reconstruction of natural feature points except for the fiducial markers. As the third step, the system constructs the landmark database including 3D positions and viewpoint dependent information of the natural feature points. We aim that these procedures in the setup phase can be done within several tens of seconds ideally, within some minutes at most.

In the MR-PreViz shooting phase, MR-PreViz movie is shot in the condition of removing the fiducial markers from the target environment. The camera position and pose is estimated initially and tracked in real-time. Here, the target environment including the lighting condition doesn't change significantly even in the outdoor environment, since the landmark database can be constructed in a short time.

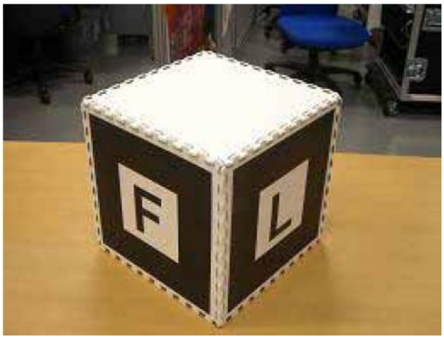


Fig. 3. Cubic marker

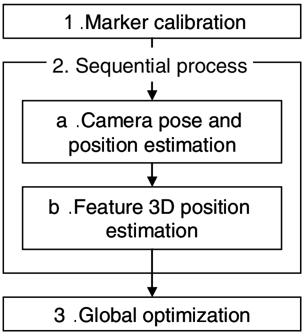

1 .Marker calibration
2. Sequential process
a .Camera pose and position estimation
b .Feature 3D position estimation
3 .Global optimization

Fig. 4. Overview of 3D reconstruction

3 Construction of the Landmark Database

In MR PreViz, the landmark database should be constructed within minimal time and efforts, since the camera work and shooting sometimes changes at the location site. This section shows a rapid construction method of landmark database suitable for MR-PreViz.

3.1 3D Reconstruction

In order to estimate 3D position of feature points using structure-from-motion, the camera position and pose is required while capturing the feature points. The proposed method uses fiducial markers arranged in the real environment to estimate the camera position and pose. We adopted the cubic markers as shown in Fig.3 which are constructed by some ARToolKit markers [7], since such cubic markers make it possible to estimate camera position and pose more accurately than the planar markers. We assume multiple cubic markers are used to realize 3D reconstruction in wide area.

The flowchart of 3D reconstruction is shown in Fig.4. At first, relative positions of cubic markers have to be calculated (marker calibration). This process can be skipped if the target environment is covered by a single marker. Second, camera parameter and the 3D position of feature points are estimated for the captured image sequences. Image features are detected by FAST corner detector [12] which is known as one of the fastest detectors but has high repeatability. After detecting and identifying markers by an ARToolKit module, the camera parameters are estimated by solving PnP problems [11]. Finally, camera parameters in all frames and the 3D positions of natural feature points are refined by minimizing the formula (1).

$$E = \sum_{p} \sum_{f} \left| \mathrm{x}_{fp} - \hat{\mathrm{x}}_{fp} \right|^2 \tag{1}$$

Here, x_{fp} is the position of the detected image feature and $\hat{\mathrm{x}}_{fp}$ is the projected position of the feature point p in the frame f . This optimization is achieved by bundle adjustment using Levenberg-Marquardt method [8, 9].

3.2 Landmark Database

The each entry of the landmark database includes the following components:

(a) 3D position
(b) Viewpoint dependent information

- SIFT feature
- Scaling factor
- Position and pose of the camera

In order to realize rapid and robust tracking, we adopt SIFT feature for the matching. The scaling factor of every viewpoint is also required since the characteristic scale of SIFT is calculated by the distance between camera position and 3D position of landmarks in the proposed method. Details are described in section 4.1.

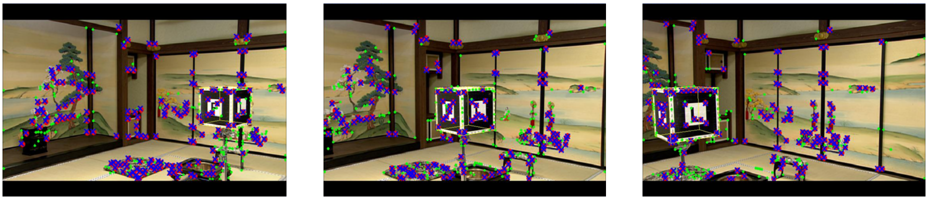
In addition to the landmark database, the key frame database is needed to estimate the initial camera parameter. The key frame database stores landmark data for each frame. Key frames are chosen automatically from captured image sequences at a regular interval. Each entry of the key frame database consists of 3D positions and SIFT features of visible landmarks.

3.3 Performance

We checked the performance of the proposed procedures of constructing the landmark database. In this experiment, we constructed the landmark database of Japanese traditional room from a 150 frames video sequence capturing a cubic marker. We used a PC (Xenon 3.4GHz, 2GB RAM) and a video camera (Sony, HDW-F900R, 720×364, 30fps) and the camera moved along the curved rail.

Fig.5 shows estimated 3D positions of tracked feature points and Fig.6 shows the estimated camera path from feature points and the cubic marker. Tab.1 shows the average processing time of every frame. Totally, it took about 27 seconds to estimate the 3D positions of all feature points.

Removing some feature points on the cubic marker from the 3D reconstruction result of the feature points, we constructed the landmark database including 288 feature points. Totally, it took about 40 seconds to construct the landmark database including the 3D reconstruction step. We can say that the proposed method for constructing the landmark database is enough fast and this method is suitable for MR-PreViz.



50th frame 100th frame 150th frame

(●detected landmarks, ×projected landmarks, ●tracked features)

Fig. 5. Estimation result of features

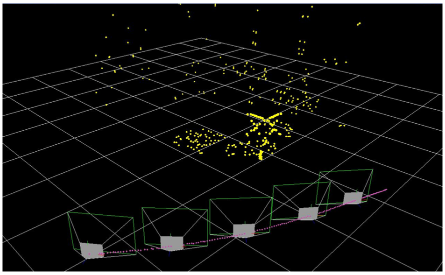


Fig. 6. Result of 3D reconstruction

Table 1. Processing time of 3D reconstruction (1 frame)

Process	Time (msec)
Camera parameter estimation	6.0
Feature 3D position estimation	45.5
Other processes	7.5
Total	59

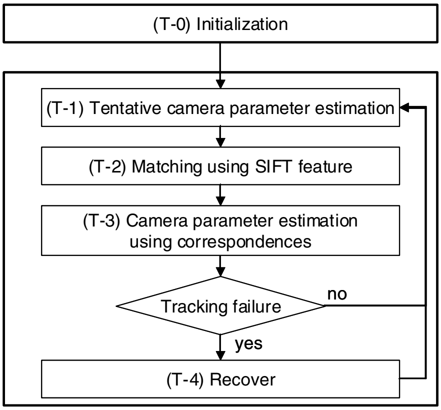

(T-0) Initialization
(T-1) Tentative camera parameter estimation
(T-2) Matching using SIFT feature
(T-3) Camera parameter estimation using correspondences
Tracking failure
no
yes
(T-4) Recover

Fig. 7. Flow chart of tracking

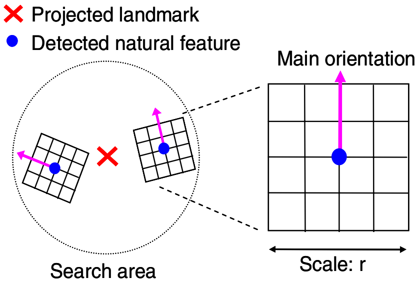


Fig. 8. SIFT descriptor layout for 4×4 sub-regions

4 Camera Tracking

In the MR-PreViz shooting step, the camera position and pose is estimated in real-time using the landmark database (Fig.2.4). Fig.7 shows the flow chart of the camera tracking. At first, the initial camera parameter is estimated by using key frame database (T-0). Next, tentative camera parameter is estimated by matching landmarks between successive frames (T-1). Patches of a size of 10×10 pixels are extracted and the patch similarity is measured by a sum of absolute difference. To determine the correspondences between landmarks and feature points in an input image, SIFT features of landmarks are matched to input image (T-2). Finally, the camera parameter is estimated from the list of 2D and 3D correspondences (T-3). If tracking fails due to some problems, a recovering process resumes tracking (T-4).

4.1 Matching Using Modified SIFT

Previous method [6] needs a large computational cost because multi-scale image templates have to be constructed for matching landmarks to detected points. In order to reduce the computational cost, our approach uses a modified SIFT descriptor [10].

Generally, SIFT is not suitable for real time operations because it takes amount of time for calculation of the scale factor. Wagner [13] succeeds reducing computational cost by fixing the scale, but this method does not have the scale invariance. Therefore, when the distance between the camera and the feature point is changed, the feature quantity is also changed in Wagner's SIFT. Accordingly, our proposed method tries to reduce the computational cost and memory by calculating the scale from the distance D between the landmark and the camera. The layout of SIFT descriptor is shown Fig.8.

SIFT features used in our proposed method (modified SIFT) is obtained by the following steps.

1. Detecting matching candidates for landmark i using FAST corner detector
2. Calculating the scale by the formula (2)

$$r_i = \frac{r'_i \times d'_i}{d_i}. \tag{2}$$

   Here, r' and d' means the scale of the feature and the distance between the landmark and the camera while used in the constructing the landmark database, respectively. d represents the distance between the tracked landmark and the camera.

3. Rotating the descriptor region toward the main orientation which is obtained by calculating the gradient orientations and magnitudes
4. Describing the normalized 128 dimensional vector

After matching landmarks to detected natural feature points, mismatched points are removed by PROSAC [14] and the final camera position and pose are estimated.

4.2 Recovering from Tracking Failure

Generally, camera tracking sometimes fails due to blurring or occlusions. Accordingly, the camera tracking method should be able to recover automatically form tracking failures. Our proposed method realizes automatic fast recovering from tracking failures on the assumption that the camera position and pose does not change significantly between before and after the tracking failure frame. Concretely, our proposed method solves the following problem.

The matching cost increases since all detected feature points from image have to be treated as matching candidates. All of them are matched to all landmarks. To solve this problem, our proposed method links feature points between before and after the tracking failure frame using the nearest neighbor search.

4.3 Initialization

In the first frame, the system has to estimate the camera position and pose without tracking techniques. The proposed method realizes the initialization on the assumption that the camera position is not so far from the camera path of the setup phase. This initialization is realized by using the key frame database.

The initialization consists of the following two steps.

1. Finding the nearest key frame

   As the first step, the system searches the nearest key frame by comparing the input image with key frame images of the key frame database. The similarity of images is given by the following formula

$$S_j = \sum_{i=1}^{L_j} \frac{1}{SSD(v_{ji}, v')}, \tag{3}$$

where L_j is the number of landmarks registered in key frame j, and v_{ji} is the SIFT feature of landmark i in registered key frame j and v' represents the feature in current frame seemed to be the nearest neighbor of v_{ji}. This operation contributes to reduce the computational time and mismatching cases in the next step.

2. Matching using the nearest neighbor search
   In the next step, the system matches landmarks of the nearest key frame to detected feature points from the input image using the nearest neighbor search.

4.4 Experiments

We had some experiments to show the effectiveness of the proposed tracking method, comparing with the previous method [6] in the computational time and robustness. In these experiments, a notebook PC (Core 2 Extreme 2.8GHz, 4GB RAM) and a video camera (Sony, DSR-PD170, 720×480, progressive scan, 15fps) are used. The landmark database were constructed by the captured image sequence consists of 400 frames. And 10 key frames were selected manually. The SIFT scale r' was 24.

First, we confirmed the proposed method can estimate the initial camera position and pose within 45 [ms] when the initial camera position is near the camera path during constructing the landmark database. The matching process in the initialization requires 1.41 [ms] averagely for every key frame. The camera positions which succeeded the initializing are shown in Fig.9.

Second, we compared the processing time of the proposed tracking method and [6]. Tab.2 shows the processing time of these methods. Tab.2 shows that the proposed method succeeds to reduce the processing time of whole processes, especially (T2) process, considerably.

Fig.10 shows the number of matched landmarks during the camera tracking using the proposed method. This chart shows the proposed method can recover the camera tracking after tracking failure frames.

Finally, Fig.11 shows example MR images based on the camera position and pose which were estimated by the proposed method. The arrows in the right figures represent the main orientation and the circle represents the described region.

Table 2. Processing time [ms]

Process	Previous method	Proposed method
Tentative camera parameter estimation (T1)	20.6	4.2
Matching using SIFT feature (T2)	35.1	3.6
Camera parameter estimation using correspondences (T3)	3.3	0.7
Total	59.0	8.5

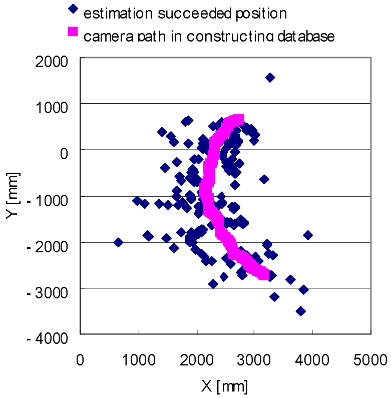


Fig. 9. Succeeded position to initialize

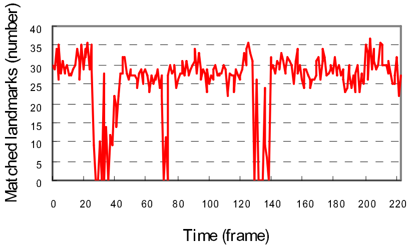
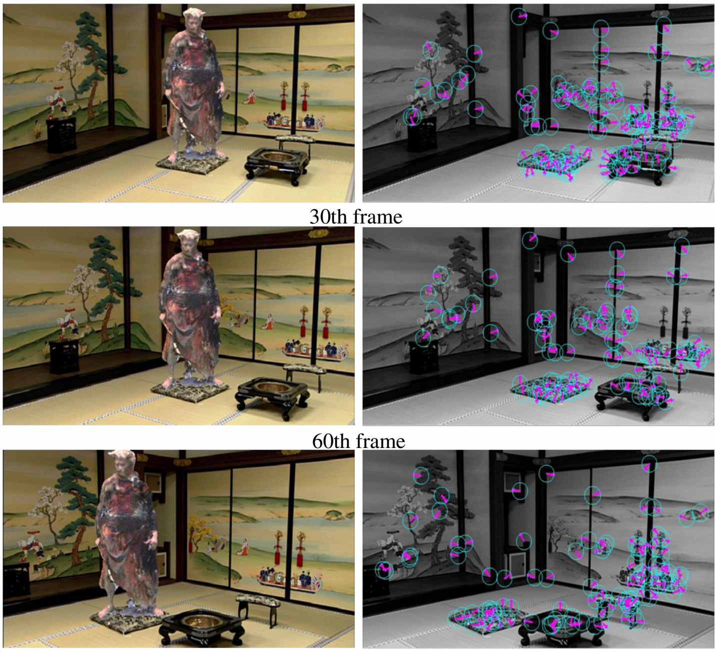


Fig. 10. The number of matched landmarks successfully



30th frame

60th frame

90th frame

Fig. 11. Left: MR images.right: Detected landmarks with SIFT

5 Conclusion

This paper describes a novel camera calibration and tracking method suitable for MR-PreViz. In the camera calibration, the proposed method tries to reduce the human and computational cost using fiducial markers. We also realized a fast and robust tracking by developing the traditional method [6] which uses the landmark database. Concretely speaking, modified SIFT and some devisal make it possible to reduce the computational costs so as to realize recovering from tracking failures.

In the future, we will shoot an MR-PreViz movie using the proposed method after polishing up the proposed method ongoingly.

Acknowledgements. This research is supported in part by Core Research for Evolutional Science and Technology (CREST) Program "Foundation of technology supporting the creation of digital media contents" of Japan Science and Technology Agency (JST).

References

1. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., Maclntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications 21(6), 34–47 (2001)
2. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: Proc. 6th Int. Symp. on Mixed and Augmented Reality (ISMAR 2007), pp. 225–234 (2007)
3. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D.: Pose tracking from natural features on mobile phones. In: Proc. 7th Int. Symp. on Mixed and Augmented Reality (ISMAR 2008), pp. 125–134 (2008)
4. Tenmoku, R., Ichikari, R., Shibata, F., Kimura, A., Tamura, H.: Design and prototype implementation of MR pre-visualization workflow. In: DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking, pp. 1–7 (2006)
5. Ichikari, R., Tenmoku, R., Shibata, F., Ohshima, T., Tamura, H.: Mixed reality pre-visualization for filmmaking: On-set camera-work authoring and action rehearsal. The International Journal of Virtual Reality 7(4), 25–32 (2008)
6. Taketomi, T., Sato, T., Yokoya, N.: Real-time camera position and posture estimation using a feature landmark database with priorities. In: CD-ROM Proc. 19th IAPR Int. Conf. on Pattern Recognition, ICPR 2008 (2008)
7. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proc. 2nd Int. Workshop on Augmented Reality (IWAR 1999), pp. 85–94 (1999)
8. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle adjustment – a modern synthesis. In: Proc. ICCV Workshop on Vision Algorithms, pp. 298–372 (1999)
9. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
10. Lowe, D.: Distinctive image features from scale–invariant keypoints. Int J. Comput.Vision 60(2), 91–100 (2004)
11. Moreno-Noguer, F., Lepetit, V., Fua, P.: Accurate Non-Iterative O(n) Solution to the PnP Problem. In: Proc. 11th Int. Conf. on Computer Vision, pp. 1–8 (2007)

12. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
13. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D.: Pose tracking from natural features on mobile phones. In: Proc. 7th Int. Symp. on Mixed and Augmented Reality (ISMAR 2008), pp. 125–134 (2008)
14. Chum, O., Matas, J.: Matching with PROSAC – progressive sample consensus. In: Proc. IEEE Compt. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 220–226 (2005)