Photo-Shoot Localization of a Mobile Camera Based on Registered Frame Data of Virtualized Reality Models

Koji Makita* AIST, Japan
Jun Nishida ZENRIN DataCom, Japan
Tomoya Ishikawa AIST, Japan
Takashi Okuma† AIST, Japan
Masakatsu Kourogi AIST, Japan
Thomas Vincent Université J. Fourier, France
Laurence Nigay Université J. Fourier, France
Jun Yamashita Univ. of Tsukuba, Japan
Hideaki Kuzuoka Univ. of Tsukuba, Japan
Takeshi Kurata‡ AIST, Japan

Abstract

This paper presents a study of a method for estimating the position and orientation of a photo-shoot in indoor environments for augmented reality applications. Our proposed localization method is based on registered frame data of virtualized reality models, which are photos with known photo-shoot positions and orientations, and depth data. Because registered frame data are secondary product of modeling process, additional works are not necessary to create registered frame data especially for the localization. In the method, a photo taken by a mobile camera is compared to registered frame data for the localization. Since registered frame data are linked with photo-shoot position, orientation, and depth data, 3D coordinates of each pixel on the photo of registered frame data is available. We conducted experiments with employing five techniques of the estimation for comparative evaluations.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Sensor Fusion, Tracking

1 Introduction

Augmented reality applications need to estimate the position and orientation of the application user or the camera to render virtual objects. Image-based tracking can realize highly accurate estimations of the camera position and orientation in real time [1]. Image-based tracking system usually create own temporal coordinate system, and be used for AR applications work in narrow area (e.g. desktop AR etc.). Moreover, combination of image-based tracking and (re-)initialization can realize AR applications work in wide area (e.g. human navigation etc.). With regard to indoor positioning, techniques based on Wi-Fi or RFID and IrDA (infrared data communication) have been developed [8] [9]. Many of these techniques are in practical use, but as they require a physical infrastructure (e.g., tags, base stations), an infrastructure database (e.g., infrastructure arrangement, radio field intensity), or both, there is the issue of how to reduce installation and maintenance costs.

In this paper, we present a method for estimating the position and orientation of a photo-shoot using virtualized reality models, as one of an initialization method for indoor environments. Figure 1 shows an outline of the proposed method. This study uses registered frame data of virtualized reality models, which are photos with known photo-shoot positions and orientations, and depth data.

*e-mail: k.makita@aist.go.jp
†e-mail: takashi-okuma@aist.go.jp
‡e-mail: t.kurata@aist.go.jp

IEEE International Symposium on Mixed and Augmented Reality 2013
Science and Technology Proceedings
1 - 4 October 2013, Adelaide, SA, Australia

Figure 1: Outline of the proposed method.

Because registered frame data are secondary product of modeling process, additional works are not necessary for creating registered frame data. In the proposed method, a photo of real environment taken by a mobile camera is compared to registered frame data. For the estimation procedure, the greatest benefit of registered frame data is that 3D coordinates of each pixel on the photo of registered frame data is available.

2 Camera position/orientation estimation using virtualized reality models

As a first step of this study, we apply three different techniques to estimate camera position/orientation, namely a technique using feature points, a technique using edge information, and a technique using surface information. A detailed description of these techniques is given below.

2.1 Technique using feature points

Firstly, registered frame data are prepared. Next, the feature points from both the input image and registered frame data are extracted and matched, and a set of 2D coordinates for the feature points of the input images and 3D coordinates for the feature points of the images that have been matched to those points is obtained. Lastly, the camera position/orientation is estimated using robust estimation with RANSAC [10] from a number of 2D and 3D coordinate sets. The techniques used for the detection of feature points are SURF (Speeded Up Robust Features) [2], ORB (Oriented FAST and Rotated BRIEF) [3], and FREAK (Fast Retina Keypoint) [4].

2.2 Technique using edge information

The virtualized reality model used in this study is made up from textured polygons. Each polygon edge can be assigned the property of being a structure edge. An example of a structure edge is shown in Figure 2. In this method, edges are first detected in the input images, and structure edges generated from the model are obtained. Multiple camera parameters for the projection of the edges are then generated, and the structure edges obtained from the model are projected onto the input images. Finally, the errors between the detected edges and the projected edges are evaluated with the respective camera parameters, and the smallest projection error above

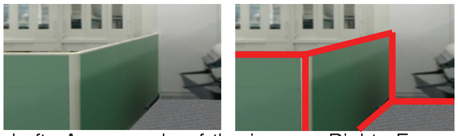

Figure 2: Left: An example of the image. Right: Examples of the edges on the structure.

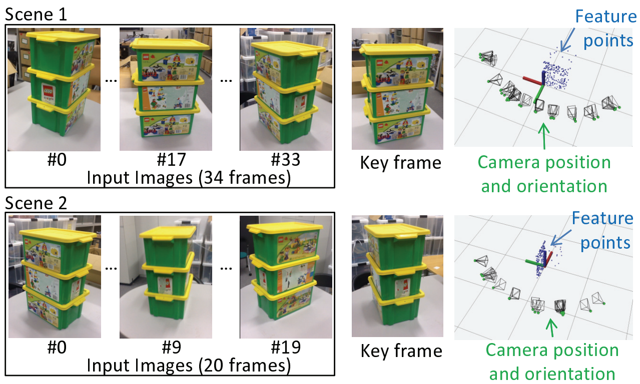

Figure 3: Input images of each scene and reference values.

the threshold is taken as the camera position/orientation estimation result. We use Canny Edge Extractor [5] for the detection of edges in the input images, and the evaluation values are obtained through the Tukey method [7].

2.3 Technique using surface information

This study takes the mutual information evaluation values [6] as surface information. First, multiple camera parameters are generated, and mutual information estimation values are calculated using input images and the template images projected with the respective camera parameters. Then, the camera parameters giving the highest estimation value are used as the position/orientation estimation result.

3 Experiment

Two scenes were prepared in the experiments, and the camera position/orientation was estimated using the input images and registered frame data for each of these scenes. Examples of the images included in these scenes are shown in Figure 3. In the following, we use “key frame” as images used for creating registered frame data. Images shot with an Apple iPad2 (resolution : 768×1024 pixels) were used in a reduced format of 180×240 pixels to shorten the processing time. For scene 1, we used 34 images shot around a stack of plastic boxes, and for scene 2, we used 20 images of the plastic boxes taken from different angles. Registered frame data for scene 1 were created using the images nearest the camera parameters for the tenth frame, and for scene 2, the images nearest the camera parameters for the seventh frame were used. We created reference values of input images with manually-positioned corresponding points.

We compared the estimated camera position/orientation and the reference values, and evaluated virtual object projection errors [11] shown in Figure 4. In scene 2, the number of frames where estimation was successful for techniques using feature points was 5 for SURF, 1 for ORB, and 0 for FREAK. The projection errors remain almost unchanged with SURF. When using edge information and mutual information, position errors and orientation errors decrease near frames 10-15 and increase further away. This may be because the camera parameter reference position at the time of the template or edge projection is close to the camera parameters near 10th frame. On the other hand, in scene 3, estimation using feature points failed for all 20 frames because the number of the feature points of the key frame was not sufficient.

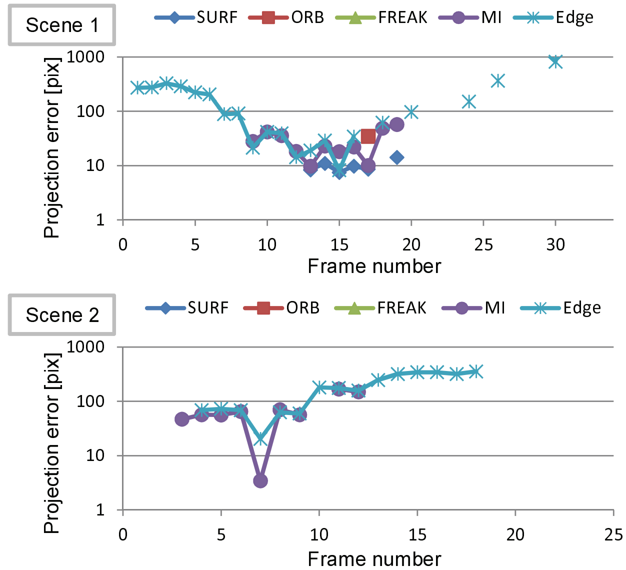

Figure 4: Projection errors.

4 Conclusion

In this paper, we described a photo-shoot localization method based on registered frame data of virtualized reality models. The results of the experiments show that, of the techniques using feature points, SURF is most suitable for estimation, and that by combining a technique using feature points with a technique using mutual information, there is the potential to expand the extent of where the camera position/orientation can be estimated.

Acknowledgements

This work was supported by Strategic Japanese-French Cooperative Program on Information and Communications Technology Including Computer Sciences (ANR and JST).

References

[1] X. Yang and K.T. Cheng: “ LDB: An Ultra-Fast Feature for Scalable Augmented Reality on Mobile Devices, ” Proc. ISMAR2012, pp. 49-57, 2012.

[2] H. Bay, T. Tuytelaars, and L. V. Gool: “ SURF: Speeded Up Robust Features, ” Proc. ECCV2006, Springer LNCS volume 3951, part 1, pp. 404-417, 2006.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski: “ ORB: an Efficient Alternative to SIFT or SURF, ” Proc. ICCV’11.

[4] A. Alahi, R. Ortiz and P. Vandergheynst: “ FREAK: Fast Retina Keypoint, ” Proc. CVPR2012, pp. 510-517, 2012.

[5] J. Canny: “A computational approach to edge detection,” IEEE Trans. on PAMI, vol. 8, no. 6, pp. 679-698, 1986.

[6] A. Dame, E. Marchand. “ Accurate real-time tracking using mutual information, ” Proc. ISMAR2010, pp. 47-56, 2010.

[7] L. Vacchetti, V. Lepetit, P. Fua, “ Combining Edge and Texture Information for Real-Time Accurate 3D Camera Tracking, ” Proc. ISMAR2004, pp.48-57, 2004.

[8] H. Yoshida, S. Ito and N. Kawaguchi: “Evaluation of Pre-Acquisition Methods for Position Estimation System using Wireless LAN, ” Proc. ICMU2006, pp. 148-155, 2006.

[9] R. Tenmoku, M. Kanbara, and N. Yokoya: “A wearable augmented reality system using positioning infrastructures and a pedometer,” Proc. ISCW2003, pp. 110-117, 2003.

[10] M. A. Fischler, and R. C. Bolles: “ A paradigm for model fitting with applications to image analysis and automated cartography, ” Comm. of the ACM, Vol. 24, pp. 381-395, 1981.

[11] K. Makita, T. Okuma, T. Ishikawa, and T. Kurata: “A benchmark suite with virtualized reality models for supporting evaluation and data set generation, ” Proc. TrakMark2012, 4 pages, 2012.