ICAT 2011, Osaka, Japan、pp.137 (2011)

# Feasibility Study on the Estimation of Photo Shoot Position and Direction Based on Virtualized Reality Environment Models

Koji MAKITA* AIST†
Jun NISHIDA University of Tsukuba, Japan
Tomoya ISHIKAWA AIST
Takashi OKUMA AIST
Jun YAMASHITA University of Tsukuba, Japan
Hideaki KUZUOKA University of Tsukuba, Japan
Takeshi KURATA AIST / University of Tsukuba, Japan

**ABSTRACT**

This paper provides a feasibility study on the estimation of photo shoot positions and directions in a modeled environment for augmented reality applications. This study focuses on the ability to compare a photo and images generated from virtualized reality environment models.

**Keywords** : Localization, Virtualized reality environment model, Model based matching, Mobile augmented reality

## 1 INTRODUCTION

Mobile augmented reality (AR) has possibility to be used in various places, conditions, and contexts. But it is difficult to achieve robust and global localization in wide areas in a single method, because each method has its own merits and demerits. Therefore, a combination of pedestrian dead reckoning and initialization methods has been attracting attention as one of the localization method in reasonable accuracy and cost [1]. For efficiently constructing an initialization method, we focus on the estimation of photo shoot position and direction by comparing the photo and images generated from virtualized reality environment models. In this study, a coarse-to-fine framework is introduced for reducing computational cost. In this framework, we evaluated three types of image similarities and an interest point matching.

## 2 ESTIMATION PROCEDURE

The coarse estimation phase (A-1,2,3) is introduced in order to select some adequate images. In the fine estimation phase (B-1,2), photo shoot position and direction are estimated. Detailed descriptions of the two phases are given below.

**(A-1) Generation of generated images from models**

Images are generated from virtualized reality environment models. Each image is generated using one camera parameter in the model's coordinate system. In this study, discretely distributed positions and directions are introduced to determine the camera parameters.

**(A-2) Calculation of similarities**

Similarities between a photo and generated images are calculated. In this study, three different types of similarities are introduced for comparison purposes. They are based on the correlation of hue-saturation histogram, SSD (Sum of Squared Differences), and ZNCC (Zero-mean Normalized Cross-Correlation).

**(A-3) Selection of generated images for fine estimation**

Some images that have high similarity are selected to be used in the fine estimation phase.

**(B-1) Matching of corresponding feature**

The photo and each generated image selected in **(A-3)** are matched according to corresponding interest points. The SIFT [2] is introduced as a method to detect interest points in the photo and generated images.

**(B-2) Estimation of photo shoot position and direction**

The photo shoot position and direction are estimated. The algorithm must be developed considering characteristics of the matching result of **(B-1)**. Therefore, the development of the algorithm is treated as a future work in this study.

*e-mail:k.makita@aist.go.jp

†Center for Service Research, National Institute of Advanced Industrial Science and Technology, Japan

## 3 EXPERIMENTAL RESULTS

Experiments were conducted at three points (Point A, B, C) in an indoor office, and photos taken at three points by an iPhone 4 phone were used. For each photo, approximately the same image (corresponding image) was generated using the corresponding camera parameter in the model's coordinate system. Figure 1 shows an example of a photo and corresponding image. Next, to generate images from models, discretely distributed positions and directions were introduced. To set the position, the height of the camera is fixed and positions are discretely set every one meter on a 2D plane. To set the direction, the roll angle and the pitch angle are fixed and yaw angles are discretely set every 5 degrees.

For the coarse estimation phase, we set maximum of the threshold by which corresponding images are included in selected images. In hue-saturation histogram based similarity, the percentage of the narrow range at point A is 43%, at point B is 69%, and at point C is 40%. In SSD-based similarity, the percentage at point A is 24%, at point B is 67%, and at point C is 97%. In ZNCC-based similarity, the percentage at point A is 10%, at point B is 7%, and at point C is 96%. In the result, ZNCC based method was dominant at point A and B. The main reason why a large region was selected at point C is a reflection of window glass. In future, characteristic features of models should be considered for similarity calculations.

For the fine estimation phase, we counted the number of correct matching of interest points. In the result, the maximum number of correct matching of interest points was 14. Therefore, new methods to determine correctness of matching must be introduced in future.

Figure 1: An example of a photo and corresponding image.

## 4 CONCLUSION

In this study, for the feasibility study of estimation of photo shoot position and direction based on virtualized reality environment models, some experimental results are described. In the coarse estimation phase, similarities based on the correlation of hue-saturation histogram, SSD, and ZNCC have been evaluated. In the fine estimation phase, matching results of interest points have been evaluated. In future studies, we need to research characteristics of the matching of interest points to develop the algorithm for fine estimation phase. For the research, we will introduce a variety of models.

**ACKNOWLEDGEMENTS**

This work was supported by ANR in France and JST in Japan.

**REFERENCES**

[1] T. Ishikawa, M. Kourogi, and T. Kurata: Economic and Synergistic Pedestrian Tracking System with Service Cooperation for Indoor Environments, Int. Journal of Organizational and Collective Intelligence, Vol.2, No.1, pp.1–20, 2011.
[2] D. G. Lowe: Distinctive image features from scale-invariant keypoints, Journal of Computer Vision, 60, 2 , pp. 91–110, 2004.