REAL-TIME OUTDOOR PRE-VISUALIZATION METHOD FOR VIDEOGRAPHERS —REAL-TIME GEOMETRIC REGISTRATION USING POINT-BASED MODEL—

Sei IKEDA, Takafumi TAKETOMI, Bunyo OKUMURA, Tomokazu SATO,
Masayuki KANBARA, Naokazu YOKOYA and Kunihiro CHIHARA

Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, Japan

ABSTRACT

This paper describes a real-time pre-visualization method using augmented reality techniques for videographers. It enables them to test camera work without real actors in a real environment. As a substitute for real actors, virtual ones are superimposed on a live video in real-time according to a real camera motion and an illumination condition. The key technique of this method is real-time motion estimation of a camera, which can be applied to unknown complex environments including natural objects. In our method, geometric and photometric registration problems for such unknown environments are solved to realize the above visualization. A prototype system demonstrates availability of the pre-visualization method.

Index Terms— Real-time Pre-visualization, Augmented Reality, Geometric and Photometric Registration

1. INTRODUCTION

This paper describes a novel pre-visualization method applicable to the entire videography industry using MR techniques, which can merge real and virtual worlds. There are essentially three stages in the videography: pre-production, production and post-production. Realizing a system helps effective shooting in the production stage by providing pre-visualized images in the pre-production stage. In the pre-production stage, the most important thing is not quality but cost for making video contents. Several works to reduce the cost already exist on real-time pre-visualization using MR techniques.

The first is a virtual rehearsal system for camera work training [1], which is a technique closer to virtual reality rather than mixed reality. It enables users to test camera work in completely virtual environments rendered by computer graphics according to a real camera motion. User can watch the current camera angle with a small monitor attached on the camera body. The camera motion is measured by a marker-based infrared tracking system. This technique can reduce the cost of stage decoration and actors.

The second one is a pre-visualization method using augmented virtuality [2]. It is designed to work in a real studio environment using a chroma key technique. Chroma key technique is extended by using a special retro-reflective cloth to provide director and actors with different views. This method enables us to reduce the cost of stage decoration.

Fig. 1. A pre-visualized scene.

The last one is a pre-visualization using augmented reality [3]. AR storyboard [4] provides intuitive interface for scene composition and camera pose/motion control in the real-environment. Such a technique reduces the cost of actors playing in the real scene.

All the pre-visualization techniques described above are designed for indoor scenes. In outdoor environments, it is difficult to obtain accurate camera pose if sensors and markers are not available. In the last kind of real-time pre-visualization techniques [3, 4], accurate 6 DOF camera pose must be estimated to register virtual actors/objects in each real video frame. However, there are not any real-time estimation method suitable for real-time outdoor pre-visualizations.

In this paper, we propose a real-time pre-visualization method for outdoor environments including complex scenes such as natural objects, as shown in Fig. 1. Our method is categorized as a pre-visualization technique using augmented reality. It is used in the pre-production stage to confirm and adjust camera work and position of actors. The insertion of the virtual objects into the real world is achieved by two key techniques: real-time estimation of camera work and lighting environment, rendering virtual actors according to the estimated ones. Thus, in this paper, the Section 2 describes the algorithm of real-time pose estimation of a camera by using a pre-defined 3-D environmental model. Section 3 describes an implementation of prototype pre-visualization system to render virtual actors in real-time. Finally, Section 5 summarizes the present work.

2. REAL-TIME REGISTRATION ALGORITHM

2.1. Geometric Registration

The geometric registration in the proposed method consists of two stages: construction of 3-D environmental model and estimation of camera pose. This section first describes our strategy of the geometric registration, then algorithm of each stage.

2.1.1. Design Strategy of Geometric Registration Method

The geometric registration is the basic technique that pose of a camera are estimated from each video frame in real-time. The registration for outdoor pre-visualization requires not only real-time processing and high accuracy but also the following two conditions: (i) modeling environment and estimating camera pose are separated; (ii) the method is available in a complex environment. We discuss the necessity of above conditions in the following introduction of previous works on real-time registration.

It is basically possible to estimate camera pose of each frame only from video data by several real-time registration methods [5, 6], which is called simultaneous localization and mapping (SLAM). The SLAM is similar to our approach in the sense of application of a structure-from-motion (SFM) technique, which is used to construct a 3-D model of the environment. The difference with our method is that the SFM is performed in online process, therefore SLAM does not satisfy the condition (i). This characteristic is not suitable to pre-visualization because location of virtual actors in the environment is difficult before modeling.

Many separate types of registration method have been also developed. In most of them, artificial 3-D polygons are used as environmental models [7–10]. It is assumed that the environment is combination of simple objects such as planes and straight lines. It is difficult to apply those methods to complex outdoor scenes including unknown natural objects.

On the other hand, there are some separate types of registration methods [11, 12] using a point-based 3-D model generated by SFM in offline process. These methods satisfy the two conditions described above. In order to separate two processes (modeling and estimating camera pose), appearance and illumination changes of feature points should be considered. Since robust identification methods are applied to find correspondences of feature points in these methods, real-time processing is not achieved. Therefore we improve this type of registration method to realize real-time processing.

2.1.2. Point-based Environmental Model

Point-based environmental model consists of feature points called landmarks, whose positions (1) are calculated by SFM using an omnidirectional camera [13]. Each landmark retains additional information for robust identification. The additional information (2) can be also calculated from the result of SFM. The elements of the model are listed as follows.

(1) 3-D position $\mathbf{x}_j$ of landmark j
(2) Information for identification
    (A) Position of omnidirectional camera $\mathbf{y}$
    (B) Multi-scale image template $\mathbf{I}_j(\mathbf{y})$
    (C) Normal vector of image template $\mathbf{n}_j(\mathbf{y})$
    (D) Scale of image template $s_j(\mathbf{y})$

2.1.3. Algorithm of Camera Pose Estimation

This section describes an algorithm estimating a camera pose (position $\mathbf{t}_i$ and orientation $\mathbf{r}_i$) of the current frame i using feature landmarks described above. This algorithm assumes that the initial camera pose is estimated by another method such as the simple method described in Section 3. Process for the subsequent frames is improvement of our previous work [12] using inter-frame feature matching. Detecting features from the input image and matching them with the landmark image templates, the correspondence between landmark and input image is then established. The following list shows these steps.

Step 1: find inter-frame correspondences Correspondences to inlier landmarks (described below) used for the previous frame are searched in the current frame i by template matching. The template matching is performed only at the observed feature points in a window whose center is predicted position of landmarks in the frame i. The predicted position of landmark j are calculated as projected positions of 3-D position $\mathbf{x}_j$ by using a camera pose of in the frame $(i-1)$.

Step 2: calculate tentative camera pose Camera pose is calculated from the inter-frame correspondences by robust estimation. We apply RANSAC [14] approach to solve the minimization problem of sum of re-projection errors.

Step 3: select landmarks This process selects landmarks used for the final camera pose estimation (Step 6) from the database. The selection criterions are designed so that the camera position $\mathbf{t}_i$ can be close to the position $\mathbf{y}$ where the selected landmark was captured with the omnidirectional camera. We use the camera position $\mathbf{t}_{(i-1)}$ of the previous frame in place of the current one.

Step 5: find corresponding landmarks Each landmark selected above step is compared with observed feature points by template matching. Template matching requires rectification of image template $\mathbf{I}_j(\mathbf{y})$ using its normal vector $\mathbf{n}_j(\mathbf{y})$ and scale $s_j(\mathbf{y})$. The rectification improves template matching even if two positions $\mathbf{t}_i$ and $\mathbf{y}$ are different. We also use normalized correlation in this step to improve template matching for illumination change.

Step 6: calculate final camera pose This step is basically the same as Step 2. The landmarks used for in this step are called inlier landmarks.

In Step 1 and 5, Harris operator [15] is used to detect feature points because this operator can be computed faster than more distinctive operators such as SIFT [16].

2.2. Photometric Registration

Photometric registration in general augmented reality is solved to improve quality of displayed images. However, photometric quality of presented images is not important to confirm camera work in real-time pre-visualization. In our work, the purpose of the photometric registration is representation of physical relationship between virtual object and real environment. For this purpose, important photometric property is cast shadow.

The approach of our photometric registration is based on the image based shadowing [17]. In this method, cast shadows on a known surface are simulated by using a simplified light environment acquired with a wide field-of-view camera. The flow of our method is as follows. (1) Acquired image of each frame is mapped onto a sphere surface, assuming positions of light sources are located at infinity. (2) Each frame of the acquired image is then binarized and labeled to obtain a dominant light source with maximum area Ω. (3) Direction of the dominant light source is calculated from the center of gravity of the area Ω. (4) Hard shadow by parallel lights is rendered on a flat plane by using the shadow map method [18].

3. IMPLEMENTATION OF PROTOTYPE PRE-VISUALIZATION SYSTEM

This section describes a prototyped real-time pre-visualization system. The system is designed to test performance of our method and to introduce the technique for general visitors at certain event site.

Configuration of the system is shown in Fig 2. Specifications of these devices are also shown in Table 1.The prototyped system consists of two cameras, two PCs and two displays. One of the two cameras is a handy camera used to take objective contents. Another is a fixed camera for acquisition of the light environment. Characteristic of the prototype system is that two PCs are used to divide whole the processing into visualization and estimation of camera pose. For this purpose, the captured video should be divided into two streams. Estimated camera poses are transmitted from the first PC to the second via Ethernet. The displays are used to show visualized virtual actors and result of corresponding features.

In the construction of a point-base environmental model, we used Ladybug2 (Point Grey Research, six XGA cameras, 30fps) as an omnidirectional camera. SFM was applied to 300 frames of omnidirectional video to obtain approximate 664 landmarks. In the estimation of camera pose, the initial camera pose was given by setting the handy camera on a tripod. The camera pose on the tripod was estimated from 10 correspondences between observed feature point clicked manually and landmark in advance. For the photometric registration, resolution of shadow map was 256×256 pixels.

As a result, we achieve 12.5 Hz update of position of virtual actors. Rendering rate of fighting motion of the actors was approximate 30 fps. Update rate of shadow map is approximate 10 fps. We could check the positions of virtual actors and our camera work in a real environment in real-time. However, the average re-projection error sometimes became approximate 4 pixels. That means virtual actors are superimposed with a little jitter at that time. In the pre-production stage, such jitter of this level is not major problem to test camera works.

4. SUMMERY

In this paper, we propose a novel real-time pre-visualization method for videographer. The method is applicable to outdoor scenes where there are natural objects such as trees and rocks. The key technique to satisfy the above characteristic is a geometric registration using a pre-constructed point-based 3-D model and feature point matching between the model and observed ones. We realized the proposed method as a prototype pre-visualization system which can display virtual objects superimposed into a real video at near video rate of a handy camera. The visualized images of virtual actors enabled us to confirm our camera works.

Acknowledgement

This paper is supported by SCOPE (Strategic Information and Communications R&D Promotion Programme) of Ministry of Internal Affairs and Communication Japan.

5. REFERENCES

[1] X. Gouchet, R. Quittard, and N. Serikoff, “SCP camera,” in *Proc. SIGGRAPH ’07: ACM SIGGRAPH 2007 emerging technologies*, 2007, p. 16.

[2] O. Grau, M. Price, and G. A. Thomas, “A 3D studio production system with immersive actor feedback,” in *SIGGRAPH ’03: ACM SIGGRAPH 2003 Sketches & Applications*, 2003, p. 1.

[3] G. Thomas, “Mixed reality techniques for TV and their application for on-set and pre-visualization in film production,” in *DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Flimmaking*, 2005, pp. 31–36.

[4] M. Shin, B.-S. Kim, and J. Park, “AR storyboard: An augmented reality based storyboard authoring tool,” in *Proc. 4th IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2005, pp. 98–99.

Table 1. Main hardwares for the prototype system.

hardware	brand name	specification
camera for pre-visualization	Sony, DSR-PD150	720×480, 15fps, progressive scan
camera for light source	ARGO Lu-135c	1024×768, 15fps, progressive scan
pc for geometric registration	Faith progress UG	CPU:Intel Core 2 Extreme 2.9Ghz, RAM:2GB
pc for rendering	Toshiba Qosmio G30/97A	CPU:Intel Core 2 Duo 2GHz, RAM:2GB NVIDIA GeForce Go 7600

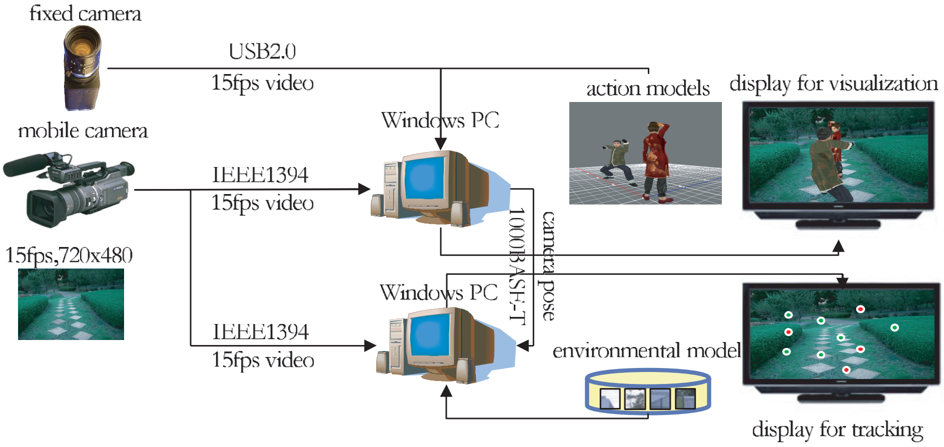

Fig. 2. Configuration of prototyped pre-visualization system.

[5] A. J. Davison, “Real-time simultaneous localisation and mapping with a single camera,” in *Proc. 9th IEEE Int. Conf. on Computer Vision*, 2003, vol. 2, pp. 1403–1410.

[6] D. Nistér, O. Naroditsky, and J. Bergen, “Visual odometry,” in *Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 964–971.

[7] A. I. Comport, É. Marchand, and F. Chaumette, “A real-time tracker for markerless augmented reality,” in *Proc. 2nd IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2003, pp. 36–45.

[8] L. Vacchetti, V. Lepetit, and P. Fua, “Combining edge and texture information for real-time accurate 3D camera tracking,” in *Proc. 3rd IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2004, pp. 48–57.

[9] H. Wuest, F. Vial, and D. Stricker, “Adaptive line tracking with multiple hypotheses for augmented reality,” in *Proc. 4th IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2005, pp. 62–69.

[10] E. Rosten and T. Drummond, “Fusing points and lines for high performance tracking,” in *Proc. 10th IEEE Int. Conf. on Computer Vision*, 2005, vol. 2, pp. 1508–1515.

[11] I. Gordon and D. G. Lowe, “Scene modelling, recognition and tracking with invariant image features,” in *Proc. 3rd IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2004, pp. 110–119.

[12] M. Oe, T. Sato, and N. Yokoya, “Estimating camera position and posture by using feature landmark database,” in *Proc. 14th Scandinavian Conf. on Image Analysis*, 2005, pp. 171–181.

[13] T. Sato, S. Ikeda, and N. Yokoya, “Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system,” in *Proc. European Conf. on Computer Vision*, 2004, vol. 2, pp. 326–340.

[14] M. A. Fischler and R. C. Bolles, “Random sample consensus: A pradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[15] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 147–151.

[16] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[17] P. Supan, I. Stuppacher, and M. Haller, “Image based shadowing in real-time augmented reality,” *Int. Journal of Virtual Reality*, vol. 5, no. 3, pp. 1–7, 2006.

[18] L. Williams, “Casting curved shadows on curved surfaces,” in *Proc. of SIGGRAPH’78*, 1978, pp. 270–274.