Computer Vision Technology Applied to MR-Based Pre-visualization in Filmmaking

Hideyuki Tamura[1], Takashi Matsuyama[2], Naokazu Yokoya[3], Ryosuke Ichikari[1], Shohei Nobuhara[2], Tomokazu Sato[3]

[1] College of Information Scienece and Engineering, Ritsumeikan University
[2] Graduate School of Informatics, Kyoto University
[3] Graduate School of Information Science, Nara Institute of Science and Technology

Abstract. In this talk, we introduce the outline of " The MR-PreViz Project " performed in Japan. In the pre-production process of filmmaking, PreViz, pre-visualizing the desired scene by CGI, is used as a new technique. In its advanced approach, we propose MR-PreViz to utilize mixed reality technology as in current PreViz. MR-PreViz makes it possible to merge the real background and the computer-generated humans and creatures in an open set or at an outdoor location. Computer vision technologies are required for many aspects of MR-PreViz. For capturing an actor's action, we applied 3D Video, which is a technology that allows one to reconstruct an image seen from any viewpoint in real time from video images taken by multiple cameras. As the other application of CV, we developed a vision based camera tracking method. The method collects environmental information required for tracking efficiently using a structure-from-motion technique before the shooting. Additionally, we developed a relighting technique for lighting design of MR-PreViz movie.

1 Introduction

Mixed reality (MR) which merges real and virtual worlds in real-time, is an advanced form of virtual reality (VR) [1]. The word "augmented reality" (AR) has the same meaning as MR. In AR space, the real world is dominant, and it is electronically daugmented and enhanced. On the other hand, MR is based on the concept of fusing the real world and virtual world by almost treating them equally. In terms of visual expression, VR deals with completely Computer-Generated Images (CGI). By contrast, AR/MR superimposes the CGI onto real scenes. Therefore, capturing the elements and analyzing and understanding attributes of the real world are necessary for AR/MR. Consequently, VR requires computer graphics technology; computer vision (CV) plays an important role for AR/MR.

AR/MR has a variety of applications. it already has been applied to medicine and welfare, architecture and urban planning, industrial design and manufacturing, art and entertainment, etc. This paper describes an application of MR technology for filmmaking, particularly the pre-visualization process. In the post-production stage of feature films, visual effects, or composing the CGIs with

Fig. 1. Conceptual image of MR-PreViz

live action images, is used routinely. Since this is operated as an off-line procedure, redoing and time-consuming processes are allowed. On the other hand, pre-visualization using MR technology in the pre-production stage requires real-time interactive merging of live actions and CGIs. This is a very difficult and challenging topic. In hopes of obtaining many fruits from this challenging theme, we are promoting the "CREST/MR-PreViz Research Project" [2][3]. Fig.1 shows the conceptual image of the MR-PreViz project.

AR/MR is a newly emerged attractive application field for CV technology. At the same time, filmmaking is a worthwhile application field to tackle. The theme "MR-based Pre-visualization in Filmmaking" makes stages for CV technology, we will introduce examples of 3 of these stages in this paper.

2 Overview of the CREST/MR-PreViz project

2.1 Significance of MR technology for PreViz

Recently, pre-visualization (PreViz, also described as "PreVis" or "pre-vis") has been used to further develop a storyboard. PreViz is a technique based on computer-generated images for visualizing action scenes, camera angles, camera blockings, lighting conditions, and other situations and conditions before the actual shoot. Compared with the conventional PreViz, which previsualizes the desired movie scene with only CGI, our MR-PreViz has significant differences:

1. MR-PreViz utilizes real backgrounds, such as sound stages, open sets, and location sites. CG objects imitate actors or creatures such as dinosaurs and aliens, which are superimposed onto the background. In terms of VR, this is an MR composite from a camera view point. In terms of filmmaking, this is on-site real-time 3D match moving.
2. Compared with virtual studios used in TV productions, MR-PreViz is a generic form which can be used in outdoor environment.
3. MR-PreViz can be used in multiple stages of PreViz from Pitch-Vis to Post-Vis, and it is especially suitable for camera rehearsals and set simulations.

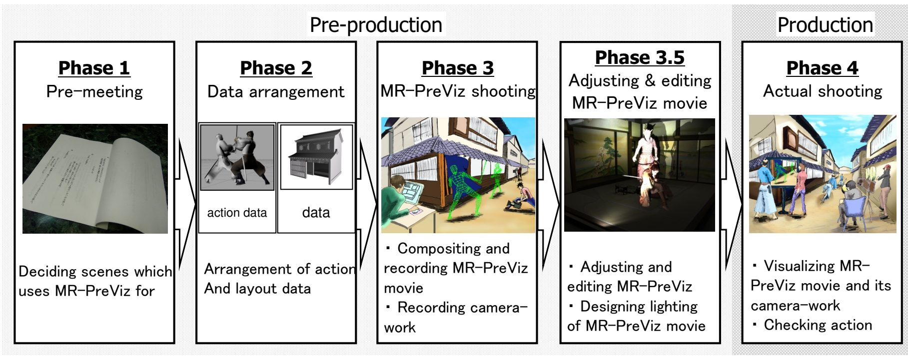

Fig. 2. Workflow of MR-PreViz

2.2 Workflow of filmmaking with MR-PreViz

The workflow of filmmaking with MR-PreViz is as follows (Fig.2).
Phase 1: Selecting scenes suitable for MR-PreViz; Scenes that should be checked using MR-PreViz, are selected.
Phase 2: Arrangement of action and layout data; We collect CG character data, animation setting data, and action data before making MR-PreViz movies.
Phase 3: MR-PreViz shooting; MR-PreViz movies are shot at the shooting site using Camera-Work Authoring Tools with a professional digital cinema camera.
Phase 3.5: Editing and Adjusting MR-PreViz movie; A high definition version of the MR-PreViz movie can be obtained by off-line rendering. Additionally, illumination on the MR-PreViz movie can be changed by relighting.
Phase 4: Application to actual shooting; The results of MR-PreViz shooting are applied to the actual shooting. Actors and staff can share ideas and images by using a MRP browser.

Recently, the processes in Phase 3.5 become more important, because the function of editing and adjusting after the MR-PreViz shooting were appreciated in experimental shootings.

3 CV technologies in MR-PreViz (1) − 3D Video

This section describes 3D video technology [4], which is used in Phase 2. While wearing a special suit has been necessary for capturing action in the past, 3D video has an advantage that the actor's actions can be captured while wearing a normal costume for the real shooting. We describe some technical highlights of our 3D video technology as computer vision research and then discuss its advantage against other possible approaches as a data source of MR-PreViz.

3.1 Introduction of 3D Video

The 3D video is media which records visual dynamic events as is. It records the actor's 3D shape, texture and motion without attaching any extra devices or markers to the object. Unlike 3D-TV, which only gives 2D stereo appearances of the scene to the human brain, 3D video explicitly estimates full 3D geometry and texture. It first captures 2D multi-view videos of the object and then estimates its 3D information purely from acquired 2D videos. Once 3D shapes are obtained from 2D videos, the original 2D images are mapped onto the 3D surface as its texture. Since this produces a conventional " 3D surface geometry + texture " style output, we can render it from an arbitrary viewpoint even with other virtual objects.

3.2 Technical Highlights of 3D Video as a Computer Vision Research

3D video technology consists of lots of challenging computer vision research topics including (1) object tracking and calibration and (2) 3D kinematic motion estimation for further editing.

(1) Object tracking and calibration: The fundamental criteria for the 3D video capture are twofold. All regions of the object surface must be observed from at least two cameras, and the intrinsic and extrinsic parameters of the cameras must be calibrated accurately. As long as we can satisfy these requirements, we can choose any combinations of cameras and their arrangement, which controls the resolution and captures area size of the system. To achieve the best combination of the resolution and capture area with a fixed image resolution, we have developed an active (pan-tilt-zoom) camera system named " cell-based 3D video capture " (Fig.3) which tracks and captures the object on a cell-to-cell basis [5]. In this approach, we can reformulate the original online tracking and calibration problem as a cell arrangement and tracking problem.

(2) 3D kinematic motion estimation: 3D video consists of a time-series of 3D surface geometry and texture information and does not have any information on the kinematic structure of the object. The goal here is to estimate the kinematic structure and posture of the object by observing its surfaces. Fig.4 shows a result for a complex posture. The key point here is explicit management of the 3D surface areas invisible from any cameras which have less accuracy on the surface geometry [6].

3.3 3D Video for MR-PreViz

The key point of 3D video technology for MR-PreViz is its geometry-based representation of the scene. This property brings (1) seamless integration with other 3D virtual objects, and (2) free-viewpoint rendering for pre-visualization. In addition, once we obtain the posture information of the captured object, we can edit them with conventional CG techniques.

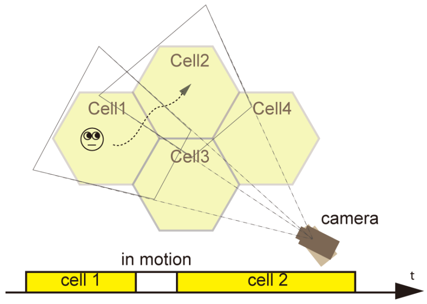

Fig. 3. Cell-based 3D video capture. The cameras cooperatively track the object on a cell-to-cell

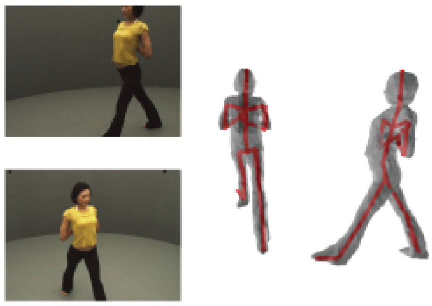

Fig. 4. Complex posture estimation

4 Camera tracking using landmark database

This section describes a camera tracking method used in Phase 3. The proposed method has a significant role to play in composing CGI onto real background in out-door environments as an on-site real-time 3D matchmove.

4.1 Registration method using landmark database

In order to overlay CG actors at geometrically correct positions in captured images, position and posture parameters of a video camera are necessary. High-accurate measurements of these parameters has been achieved by using combinations of sensors attached to the video camera and emitters arranged in the target environment, such as ultrasonic, magnetic or optical sensors/emitters. Although these methods work well in a small environment like a TV studio, it is not a realistic scenario to use them in a large outdoor environment due to the difficulty in arranging and calibrating those emitters. A combination of GPS and other sensors is one of possibilities for an outdoor environment but its accuracy has not reached the practical level of geometric registration in MR.

On the other hand, vision-based methods can estimate camera parameters without external sensors. The PTAM [7] is one of the famous methods that estimates camera parameters in real-time by tracking feature points on input images. This method obtains relative camera motion and 3-D positions of feature points simultaneously without prior knowledge. One problem in the PTAM for MR-PreViz is that absolute position and posture information have never been recovered. This limitation makes pre-arrangement of CG objects difficult.

Landmark database (LMDB) [8] is one of the promising approaches to this problem. In this method, as an offline stage, the target environment is captured by using an omnidirectional camera, and feature points tracked in this video sequence are registered to the database as landmarks. 3-D coordinate of feature

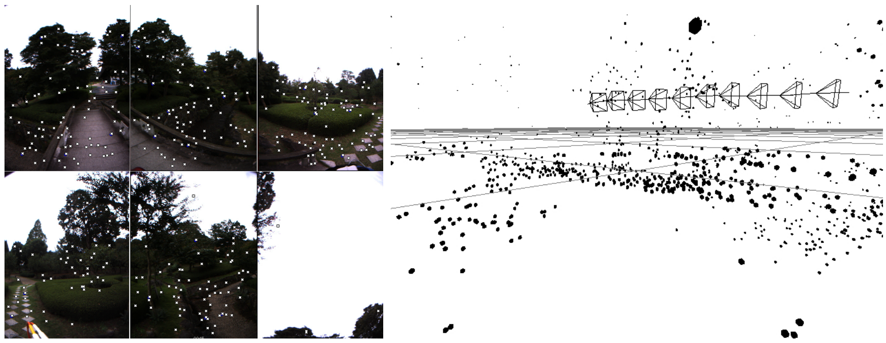

Fig. 5. Detected feature points in omnidirectional image (left) and their 3-D positions estimated by structure from motion (right).

points are estimated by using a structure-from-motion technique for omnidirectional video that simultaneously estimates 3-D positions of feature points and camera motion [9].

Fig.5 shows detected positions of feature points on the omnidirectional video sequence and estimated 3-D positions of feature points. Absolute 3-D positions of landmarks are recovered by using several reference points whose absolute 3-D positions are manually measured. Visual information of each landmark is also stored to the database from the omnidirectional video.

In the online stage, pre-registered landmarks are searched for from each image of the input video using its visual information. After finding corresponding pairs of landmarks and feature points, absolute position and posture of the video camera are estimated by minimizing re-projection errors of these pairs. By using estimated camera parameters, CG characters are rendered from the appropriate viewpoint and they are finally merged into the input image as shown in Fig.6.

4.2 Rehearsal Path Method: Refinement of the registration method using LMDB

The registration method using LMDB was originally developed for a general scenario of MR. We can utilize constraints of the usage in filmmaking for refinement [10]. In particular, it is assumed that a rough camera path is known in filmmaking. We have develop a method called "Rehearsal Path Method; RPM" which refines efficiency and accuracy of the registration method by restricting the moving range of camera to the camera path during construction of LMDB. The RPM automatically gathers information of the shooting site by pre-shooting and constructs a landmark database (LMDB). The RPM consists of two phases as shown in Fig.7.

Rehearsal Phase: In this phase, RPM utilizes a video sequence with a fiducial marker captured during the rehearsal as a learning sequence. The geometry of the site is estimated using a structure-from-motion technique. In particular, the positions of feature points in 3D space are first estimated by using epipolar

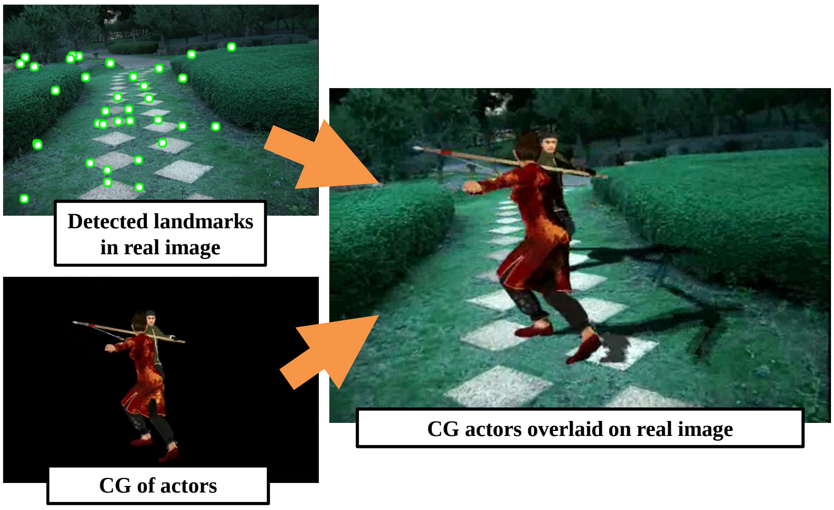

Fig. 6. Geometric registration between CG object and real scene using LMDB.

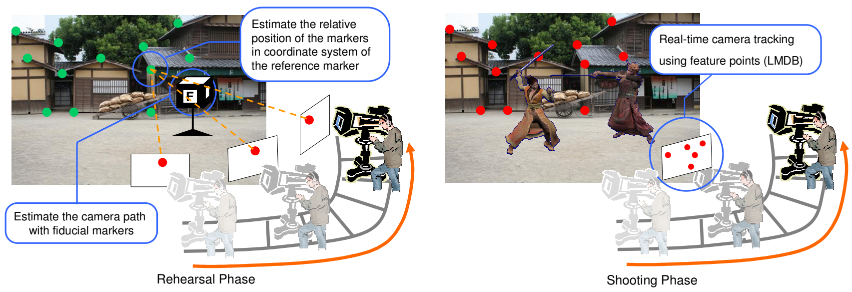

Fig. 7. Rehearsal Path Method (RPM)

geometry on several frames in the video sequence. Secondly, 6DOF parameters of the camera and positions of new feature points are simultaneously calculated by tracking the feature points. Finally, the coordinates of the 3D points are transformed into real world coordinates by recognizing the marker. SIFT, a local invariant, of each landmark is calculated and entered into our LMDB with 3D positions.

Shooting Phase: In this phase, the fiducial marker is removed for MR-PreViz shooting. The registration of virtual world and real world is realized by correlating the 2D feature points in the images with the 3D points in the LMDB. The RPM is possible to automatically obtain an initial position and recover from tracking failure by using SIFT matching between a present frame and keyframes on the camera path of the LMDB prepared in advance. The RPM successfully refined efficiency and accuracy of the registration method by using the knowledge of camera path. Besides PreViz, The RPM is also applicable to other purposes which have the same constraint.

5 CV technologies in MR-PreViz (3) - IBL and Relighting

This section describes the method assumed to be used in Phase 3.5, which is the additional process after the MR-PreViz shooting. In addition improving the speed for processing, this method also can be used for the MR-PreViz shooting in Phase 3. Photometric consistency between virtual objects and a real scene is one of the most important issues in the CG research area. "Virtual Cinematography" [11], developed by Debevec et. al., is a famous example which applied illumination technology researched in an academic field to film making. This work is based on image based lighting (IBL). IBL enables the lighting condition in the desired scene to be reconstructed by utilizing a series of images or omnidirectional images stored in advance.

Light Stage is the dedicated equipment for translating the IBL concept into reality as shown in Figure 6. Light Stage enables illumination onto actors to be changed freely by systematically illuminating structured lights. Previously, several versions of light stage have been developed. One group received the Academy Award in 2010 for their technical contributions to "Spiderman 2" and "Avatar". Light Stage was originally used in the postproduction process for keeping photometric consistency. By contrast, we are developing a visualization method for "Look", which refers to the feeling of an image provided by illumination and color tone. For visualizing the look that directors and cinematographers imagine, we developed relighting technology for MR-PreViz movies. Light Stage realizes relighting for the actor who appears on the stage. On the other hand, the target of our relighting for MR-PreViz is the background of the indoor-outdoor location. Therefore the approach using structured lights can not be used in our case.

In the MR research area, there are many works for photometric consistency between virtual objects and real objects. Lighting conditions are estimated for illumining the virtual objects on the same condition of the real scene. Our research takes a lateral approach. As a next step, we will focus on challenging trials to Look-Change of MR space. In this section, we introduce a relighting method for the Look-Change [12]. The proposed method allows an MR space to have additional virtual illumination for the Look-Change. The effects of virtual illumination are applied to both real objects and virtual objects while keeping photometric consistency. There is a trade-off between the quality of the lighting effect and the processing time. Therefore, the challenge is to create an efficient model of the lighting condition of real scenes. Our method adopts a simple and approximate approach to realize indoor-outdoor relighting as shown in Fig.8.

Step 1: Preparing images without distinct shadows
If a distinct shadow, exists in the background images, it may cause a paradox between real and virtual shadows in the relighting process. We should prepare background images without shadows. We can use physical lighting equipment or image processing methods for removing shadows in the images.

Step 2: Adjusting color tone This process approximates environmental light. The color tone is adjusted by multiplying arbitrary values with respect to

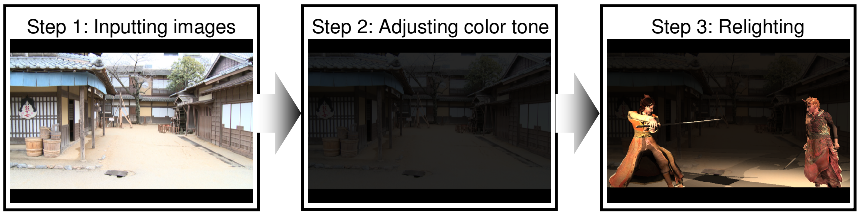

Fig. 8. Flow of our relighting method

each color channel. As a result of this process, we can change an image of daylight into an image of a night scene.

Step 3: Adding lighting effects to MR-PreViz images
After color correction, real and virtual objects in the MR-PreViz images are illuminated by virtual lighting. To optically correct illuminate objects, we estimate reflectance properties and geometry of the real objects. A relationship between pixel value and illuminance is obtained as a reflecting property. Illuminance is automatically calculated under several lighting conditions by using a reference marker where the reflecting property of the marker is known. Geometry of the site is estimated using a structure-from-motion technique.

The final MR-PreViz images of relighting are shown in Step 3 of Figure 8. Nevertheless the proposed method is developed for changing Look in a MR-PreViz movie. It is also applicable for lighting effects in MR attractions.

6 Conclusions

In this paper, we picked mixed reality as a useful target application of CV technology, and introduced mixed reality based pre-visualization in filmmaking with 3 elemental CV technologies.

These elemental technologies, as outlined below, have been steadily improved by being used many times in the actual workflow of filmmaking during the projects 5 years (since Oct 2005).

(i) The 3D video has verified its utility for the purpose of PreViz, though it does not have enough quality as a final sequence.

(ii) The camera tracking method has steadily improved its practicality by utilizing constraints of PreViz in filmmaking. Specifically, we refined the method under the assumption that a rough camera-path is decided before the shooting. We called this method "Rehearsal Path Method." The method is not omnipotent because it depends on the target scene and weather. We can improve efficiency and reliability of the method by gathering experience and setting modes based on the situations.

(iii) At first the relighting method had been used only after the MR-PreViz shooting in Phase 3.5 as an off-line process. It also can be used in Phase 3 as a real-time process.

Even conventional PreViz composed of only CG was not popular 5 years ago. However, the number of the use of PreViz in feature films was rapidly increased since then. Today, there are PreViz studios which specialize in PreViz, and "The Pre-visualization Society" has been established [13]. Although PreViz is not able to contribute to the quality of the final movie, it enables filmmakers to inspire their creativity by facilitating the process of trial-and-error in the pre-production or production stages of filmmaking. Additionally, PreViz is able to contribute to reducing the total production costs. The processes of the PreViz are subdivided into pitch-vis, tech-vis, post-vis, etc. Accordingly, MR-PreViz continues to receive much attention and makes progress for CV technology.

References

1. H.Tamura, H.Yamamoto, A.Katayama: Mixed reality: Future dreams seen at the border between real and virtual worlds. IEEE Computer Graphics & Applications **21** (2001) 64 – 70
2. R.Tenmoku, R.Ichikari, F.Shibata, A.Kimura, H.Tamura: Design and prototype implementation of MR pre-visualization workflow. In: DVD-ROM Proc. Int.Workshop on Mixed Reality Technology for Filmmaking. (2006) 26 – 30
3. R.Ichikari, R.Tenmoku, F.Shibata, T.Ohshima, H.Tamura: Mixed reality pre-visualization for filmmaking: On-set camera-work authoring and action rehearsal. Int. J. Virtual Reality **7** (2008) 25 – 32
4. J.Starck, A.Maki, S.Nobuhara, A.Hilton, T.Matsuyama: The multiple-camera 3-D production studio. IEEE Trans. on Circuits and Systems for Video Technology **19** (2009) 856 – 869
5. T.Yamaguchi, S.Nobuhara, T.Matsuyama: Cell-based object tracking method for 3D shape reconstruction using multi-viewpoint active cameras. In: IEEE Int. Workshop on Visual Surveillance (VS2009). (2009)
6. T.Mukasa, A.Miyamoto, S.Nobuhara, A.Maki, T.Matsuyama: Complex human motion estimation using visibility. In: IEEE Int. Conf. Series on Automatic Face and Gesture Recognition. (2008)
7. G.Klein, D.Murray: Parallel tracking and ma pping for small AR workspaces. In: Proc. ISMAR. (2007) 225–234
8. T.Taketomi, T.Sato, N.Yokoya: Real-time geometric registration using feature landmark database for augmented reality applications. In: Proc. SPIE. (2009) 723804–723804–9
9. T.Sato, S.Ikeda, 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. Volume 2. (2004) 326–340
10. W.Toishita, Y.Momoda, R.Tenmoku, F.Shibata, H.Tamura, T. Taketomi, T.Sato, N.Yokoya: A novel approach to on-site camera calibration and tracking for MR pre-visualization procedure. In: Proc. Human-Computer Interaction International (HCI2009). (2009) 492 – 502
11. P. Debevec: Virtual cinematography: Relighting through computation. IEEE Computer **39** (2006) 57 – 65
12. R.Ichikari, R. Hatano, T.Ohshima, F.Shibata, H.Tamura: Designing cinematic lighting by relighting in MR-based pre-visualization. In: ACM SIGGRAPH ASIA 2009 Posters. (2009)
13. http://www.previssociety.com/.