Augmented Visualization of Reality from Collection of Images

Hideo Saito

Department of Information and Computer Science, Keio University
3-14-1 Hiyoshi Kohoku-ku, Yokohama 223-8522, Japan

Keywords: Mixed/Augmented Reality, Diminished Reality, Structure from Motion, Multi-viewpoint

ABSTRACT

In this talk, I will present a future trend of visual processing in which the amount of images captured by various devices is extremely increasing. While machine learning has made drastic improvement based on the huge amount of image data, we need to achieve the innovative visualization technology using huge amount of image data. I believe that one of such trends is augmenting visual reality based on multi-dimensional visual information analysis for providing intuitive understanding using collection of various images. I will show some preliminary example research directions based on the collection images.

1. INTRODUCTION

It has been over two decades since the technology of “free viewpoint synthesis” was proposed based on 3D shape recovery and modeling from multiple viewpoint images [1] in 3D visual processing field. During the two decades, many key technologies have been developed for improving the performance of “free viewpoint synthesis”, so that we can expect it is easy to make it practical use, such as sports visualization. However, we have already noticed that making it practical use is actually not easy task, which may require more additional innovative techniques for making people feel such that impossible can be possible.

I believe that visualizing invisible reality for human is one of indispensable technologies, which can be realized by innovating 3D visual processing. According to increasing capability of amount of images captured by various kind devices, we can achieve many innovative visualizations that people have believed as impossible.

In this article, I present some example technologies for visualizing reality that cannot directly be observed, but reconstructed from collection of images. I call such technologies as augmented visualization.

2. AUGMENTED VISUALIZATION USING MOVING CAMERA

As an example technology with recent drastic improvement, we apply structure from motion (SfM) technologies [2,3] to visualization of invisible reality by recovering 3D structure and camera poses from multiple view image sequence.

Fig. 1 shows a pipeline for visualizing a video without people [4]. Here, the background of the people is visualized using different frames capturing the scene behind the people. For visualizing the background of the scene, 3D structure of the scene should also be recovered. SfM is a significant technology for obtaining the 3D structure in those cases. People in the input video are first detected and erased from the input video, then SfM is applied to recover 3D structure of the background scene. Using the estimated poses of the camera motion, the area of the detected people is filled with images capturing behind the people.

As a related technology of SfM, visual SLAMs [5] have also been studied, which takes ordered image sequence from moving camera to simultaneously recover pose of the camera and 3D structure of the scene in real-time.

Fig. 2 shows visualization of temperature on the surface of the scene captured by a set of moving a RGB camera and a thermal camera. Based on the 3D structure recovered by DSO [6] (a SLAM algorithm) of the target scene, we can map thermal images onto the color images with collect registration in 3D scene [7].

Fig. 3 shows visualization of object recognition [8] in the scene based on simultaneous 3D structure recovery and object segmentation from RGB-D image sequence [9]. Since the objects are recognized from multiple viewpoints, we can achieve better accuracy of the recognition than object recognition based on a fixed viewpoint.

3. AUGMENTED VISUALIZATION USING MULTIPLE CAMERAS

Capturing with multiple cameras also an essential way for augmented visualization. As an application of multiple camera systems, we have developed diminished reality visualizations [10].

Fig. 4 shows an example of the diminished reality visualization for baseball watching [11]. For visualizing the performance of pitcher and the ball trajectory behind the catcher and the umpire, images captured by different viewpoint cameras are used for rendering the hidden scene in the target camera.

Using multiple cameras, we have developed a system for visualizing scenes occluded by hands and tools in various hand-working scenarios [12]. In this system, we employ a RGB-D camera with multiple RGB cameras, so that we can capture the 3D structure of the hidden scene in real-time accurately. Fig.5 shows example results of visualizing the hidden working area.

In such multiple camera system, it is important to accurately estimate the extrinsic parameters of the cameras, which generally needs to detect feature points from the common scene. We have proposed a method for extrinsic calibration even without commonly detected feature points by using omnidirectional cameras [13].

4. CONCLUDING COMMENTS

The number of the cameras used for various situation will drastically be increased, which means various kinds of collections of images are available in future. Applications of desired visualization based on the increasing collection of images will definitely be spreading, so we need to develop technologies for more efficient recovery and rendering of target scenes using a large set of images.

REFERENCES

[1] Takeo Kanade, Peter Rander, Sundar Vedula, Hideo Saito, Virtualized reality: Digitizing a 3D time-varying event as is and in real time, Mixed Reality, Merging Real and Virtual Worlds, pp.41-57, 1999

[2] Changchang Wu, Towards Linear-time incremental structure from motion, 3DV 2013, 2013.

[3] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Brian Curless, Steven M. Seitz and Richard Szeliski, Building Rome in a day, Communications of the ACM, Vol. 54, No. 14, pp. 105-112, October 2011.

[4] Kentaro Yagi, Kunihiro Hasegawa, Hideo Saito, Diminished reality for privacy protection by hiding pedestrians in motion image sequences using Structure from Motion, Proc. ISMAR2017 Workshop (HDCD4MAR), Oct. 2017.

[5] Takafumi Taketomi, Hideaki Uchiyama and Sei Ikeda, Visual SLAM algorithms: a survey from 2010 to 2016, IPSJ Transactions on Computer Vision and Applications, 9:16, 11 pages, 2017.

[6] J. Engel, V. Koltun, D. Cremers, Direct Sparse Odometry, In arXiv:1607.02565, 2016.

[7] Masahiro Yamaguchi, Hideo Saito and Shoji Yachida, Application of LSD-SLAM for visualization temperature in wide-area environment, Proc. the 12th International Conference on Computer Vision Theory and Applications (VISAPP2017), Vol. 4, pp.216-223, ISBN: 978-989-758-225-7, Feb. 2017.

[8] Yoshikatsu Nakajima, Shohei Mori, Hideo Saito, Semantic object selection and detection for diminished reality based on SLAM with viewpoint class, Proc. ISMAR2017 Workshop (HDCD4MAR), Oct. 2017.

[9] K. Tateno, F. Tombari, and N. Navab, Real-time and scalable incremental segmentation on dense slam, International Conference on Intelligent Robots and Systems (IROS2015), pp. 4465-4472, 2015.

[10] Shohei Mori, Sei Ikeda and Hideo Saito, A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects, IPSJ Transactions on Computer Vision and Aplications, 9:18, 2017

[11] Takanori Hashimoto, Yuko Uematsu, Hideo Saito, Generation of see-through baseball movie from multi-camera views, Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP2010), pp.432-437, 2010

[12] Shohei Mori, Momoko Maezawa and Hideo Saito, A work area visualization by multi-view camera-based diminished reality, Multimodal Technologies and Interact. 3, 18, 12 pages, 2017.

[13] Shogo Miyata, Hideo Saito, Kosuke Takahashi, Dan Mikami, Mariko Isogawa, Akira Kojima, Extrinsic camera calibration without visible corresponding points using omnidirectional cameras, IEEE Transactions on Circuits and Systems for Video Technology, published on July 2017.

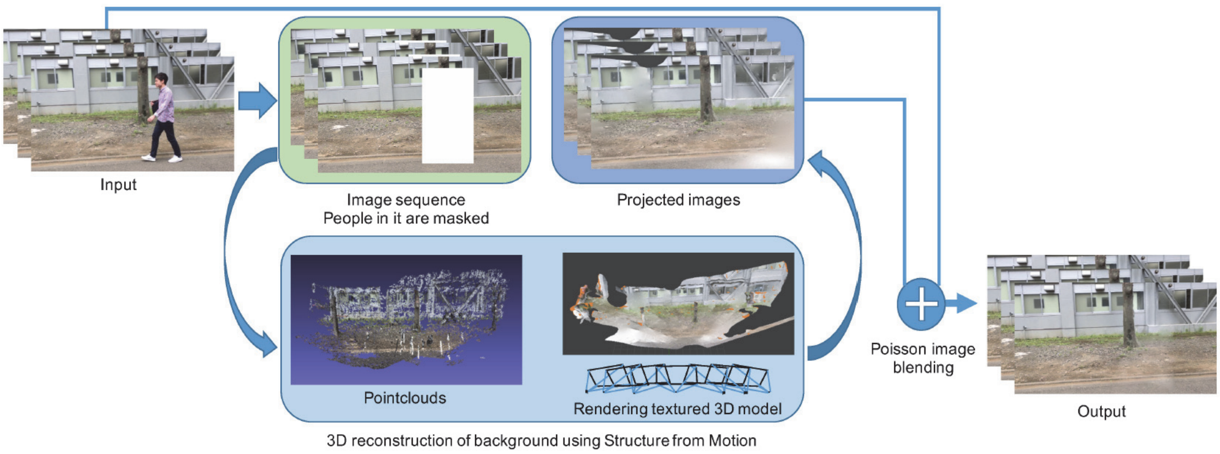
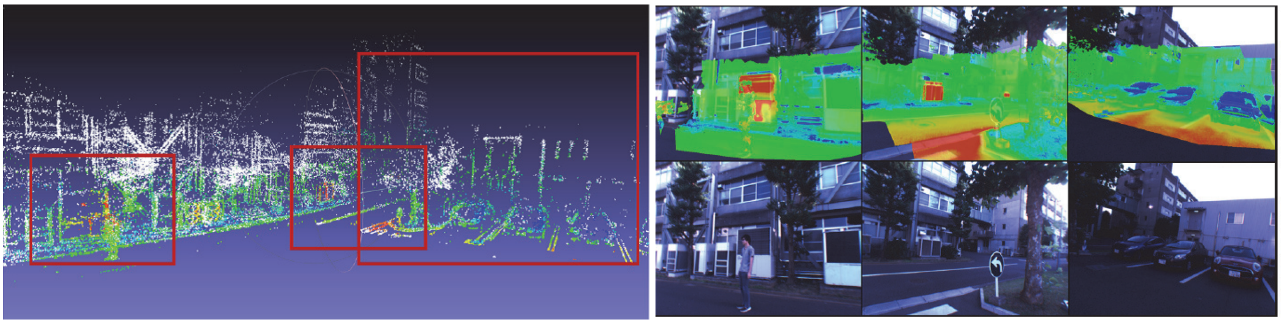


Fig.1 Pipeline for visualizing a video without people [4].



(a) 3D structure recovered by DSO with color-coded temperature

(b) Thermal distribution rendered onto RGB images.

Fig. 2 DSO (a SLAM algorithm) recovers 3D structure of the scene (a) from multiple view images. This is also used for rendering thermal distribution onto the RGB images (b) [7].

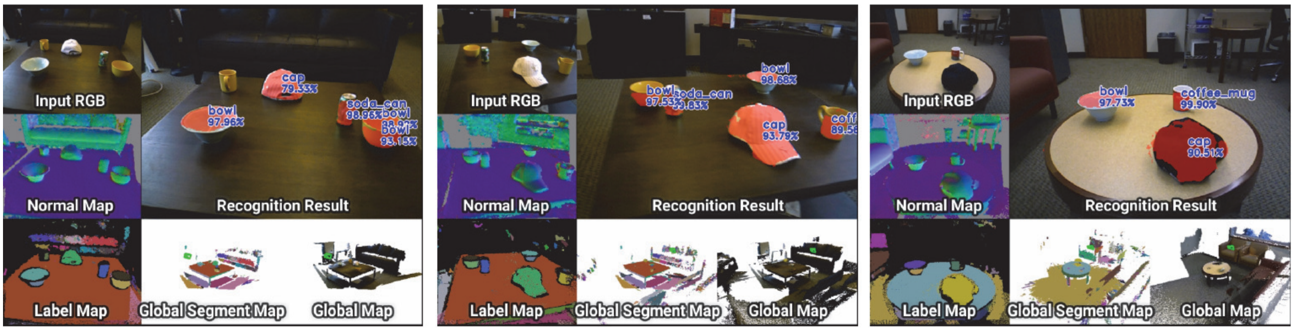
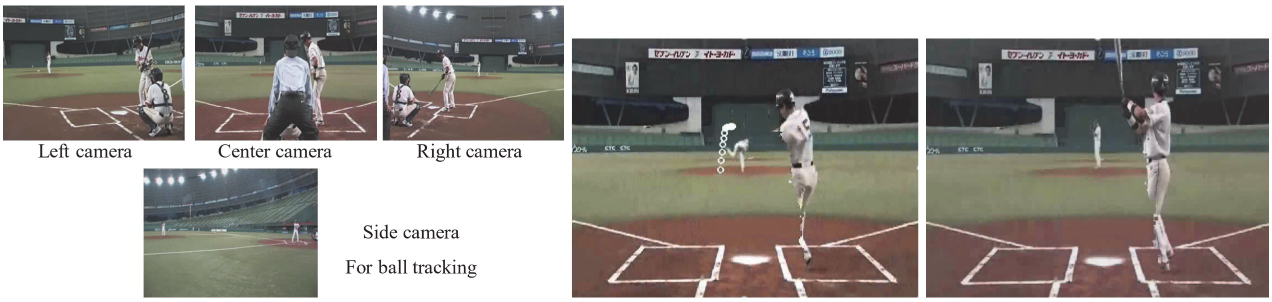


Fig. 3 Object recognition results by multi-view-based recognition method working in real-time while processing SLAM, segmentation, and object recognition [8].



(a) Input multi camera images

(b) Visualization of hidden scene by catcher and umpire

Fig. 4 Diminished reality visualization for baseball watching [11].

(a) Setup of multiple cameras

(b)Original image

(c) Hidden work area visualization

Fig. 5 Diminished reality for hidden working area visualization [12].