Display Methods of Projection Augmented Reality based on Deep Learning Pose Estimation

Hyocheol Ro
Media System Lab, Yonsei University
Seoul, Republic of Korea
hyocheol.ro@msl.yonsei.ac.kr

Yoon Jung Park
Media System Lab, Yonsei University
Seoul, Republic of Korea
yoonjung.park@msl.yonsei.ac.kr

Jung-Hyun Byun
Media System Lab, Yonsei University
Seoul, Republic of Korea
junghyun.byun@msl.yonsei.ac.kr

Tack-Don Han
Media System Lab, Yonsei University
Seoul, Republic of Korea
hantack@msl.yonsei.ac.kr

ABSTRACT

In this paper, we propose three display methods for projection-based augmented reality. In spatial augmented reality (SAR), determining where information, objects, or contents are to be displayed is a difficult and important issue. We use deep learning models to estimate user pose and suggest ways to solve the issue based on this data. Finally, each method can be appropriately applied according to various the applications and scenarios.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; *User centered design*; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Spatial Augmented Reality, Projection Augmented Reality, Mixed Reality, Deep Learning, Pose Estimation

ACM Reference Format:
Hyocheol Ro, Yoon Jung Park, Jung-Hyun Byun, and Tack-Don Han. 2019. Display Methods of Projection Augmented Reality based on Deep Learning Pose Estimation. In *Proceedings of SIGGRAPH '19 Posters.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3306214.3338608

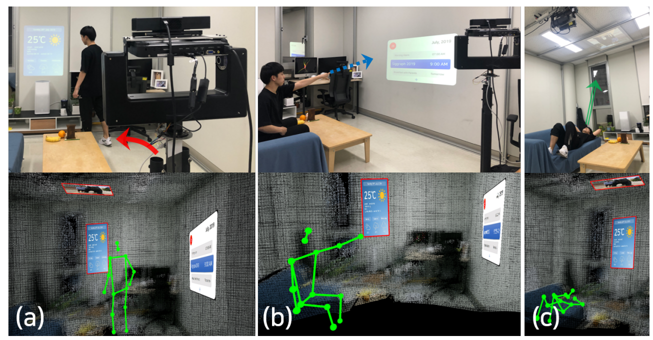

Figure 1: (a) Passing Near a Content. Red line is User Path. (b) Pointing a Surface. Blue Dotted Line is Pointing Ray-casting. (c) Looking a Head to Ceiling. Green Line is Head Direction.

1 INTRODUCTION

In spatial augmented reality (SAR), determining where information, objects, or contents is to be displayed is a difficult and important issue. In head-mounted display (HMD)-based augmented reality (AR) environments such as Microsoft's HoloLens, naturally augment window objects when a user views places where the virtual windows are already registered [Chen et al. 2015]. In a projection-based AR environment using a single projector, on the other hand, the direction in which the projected display is possible is limited to only one direction. For this reason, studies have been conducted on constructing an immersive projection room that can cover all directions in a cubic space using multi-projectors (e.g., CAVE [Bimber and Raskar 2005]). However, such a design is costly and cumbersome for users to install.

We use a projection-based AR system with a single projector, that is mounted on a pan-tilt system and can rotate 360 degrees horizontally and vertically. As a result, it is not only capable of providing projection displays in all directions, but it also has a simple and movable structure. In this system, overall geometric information about the space is acquired to reconstruct a three-dimensional map, and then contents/information are registered in advance on the detected plane. After that, we implemented a mixed reality (MR) platform by allowing the user to identify contents/information at specific locations through the augmented projection.

Unfortunately, we require the system to determine the direction in which the content should be displayed. In other words, for the task of selecting only one plane at a time, the user must manually select it or the system should automatically suggest it. In this paper, we propose a SAR-providing method based on a user location tracking technique to solve this problem.

2 RELATED WORK

Studies have variously investigated methods for determining the location where to display the projection. In *HeatSpace* [Fender et al. 2017], they tried to find the optimal surface by analyzing behavior based on geometric information of the user's space, distance and

SIGGRAPH '19 Posters, July 28 - August 01, 2019, Los Angeles, CA, USA

ACM ISBN 978-1-4503-6314-3/19/07.
https://doi.org/10.1145/3306214.3338608

visibility and generating it as a heatmap. Meanwhile, in a study by Shafaq Mussadiq *et al.* [Mussadiq et al. 2017], they used a binary code pattern to find a sub-region where no occlusion or overshadow occurred, allowing the projection to be confined. However, these studies focused on how well the homography is displayed on the sub-surface before registering new content. When contents are already registered on the surface in the MR space, determining the direction that the user wants to display contents has not been addressed.

3 SYSTEM CONFIGURATION

In recent years, recognition technology using deep learning has been greatly developed in the field of computer vision. Among them, pose estimation that can grasp the user's joints has also been studied. Early deep learning models had high recognition performance but also had drawbacks of slow speeds, making them virtually impossible to use in real time. However, many studies have recently overcame this problem, and high-performance and high-efficiency recognition has become possible. In addition, the recent computer vision research is an open source environment in which the source code is released to the public.

Using deep learning models, we implemented a prototype that can accurately grasp the user's position in an AR/MR environment and provide a projection in the proper direction. Here, not only body tracking but also facial landmark technology were applied to solve the problem that cannot be distinguished from the front and rear of the user.

In order to build a projection-based AR platform that covers 360 degrees in a smart home environment, it is important to acquire the geometric information around the system first. We rotated the RGB-D camera with the mounted pan-tilt, reconstructed the 3D map information around the user and then detected the planes. We have implemented scenarios that can augment contents/information based on the user's location or viewing direction. There are three main methods to display the contents/information by designating the projection area naturally depending on the application.

One scenario is when a user passes near a plane where certain content is registered. This is useful when a user wants to interact with an object at a specific location and augment related information [Ro et al. 2019]. A second scenario is when the system must project the contents in the direction the user wants. It will be possible to use it in applications that require more active user interaction. A third scenario is when the user is looking straight ahead, and the projector must display the content in that direction. For example, we can imagine scenarios where the user is sitting on a couch, lying in bed, or sitting at a desk [Ro et al. 2018]. In all three cases, it is important to accurately estimate the user's pose. The kind of data required is somewhat different depending on the case. In the case of projection around a user, only the location information of the user is required. On the other hand, when projecting to the direction that the user views, the user's position and major joint data related to the orientation are required. Here, the user's facial direction through landmark detection can also be used as an auxiliary. Finally, when the user manually sets the display direction, data information is required for the user's arm joints.

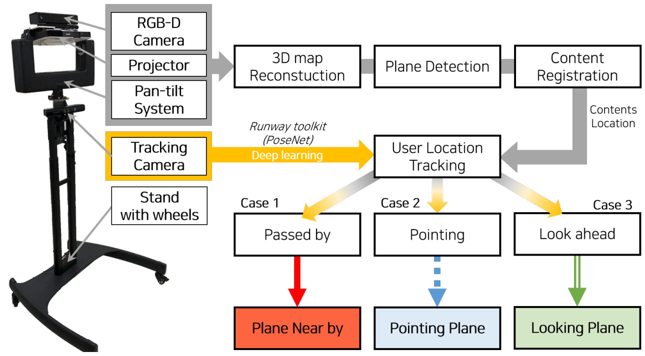

Figure 2: System Configuration and Flow Chart.

4 CONCLUSION AND FUTURE WORK

This paper describes the implementation of the three main projection-based AR display-providing methods that can come from real-world scenarios. Each method can be applied differently depending on the application or user situation, but the underlying technology is eventually a pose estimation that traces the user's location. We were also able to easily prototype the open source model *PoseNet* model [Papandreou et al. 2017] and the *Runway* toolkit [run 2019]. Of course, further work needs to be done in order for these projection-direction methods to be applied in real applications. In particular, it is essential to test which of the three methods is the most intuitive and convenient for the user. It should also be noted that these methods may have different results for different applications and scenarios.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No.NRF-2018R1A2A1A05078628).

REFERENCES

2019. *Runway*. https://runwayapp.ai/
Oliver Bimber and Ramesh Raskar. 2005. *Spatial augmented reality: merging real and virtual worlds.* AK Peters/CRC Press.
Henry Chen, Austin S Lee, Mark Swift, and John C Tang. 2015. 3D collaboration method over HoloLensâĎć and SkypeâĎć end points. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences.* ACM, 27–30.
Andreas Fender, David Lindlbauer, Philipp Herholz, Marc Alexa, and Jörg Müller. 2017. HeatSpace: Automatic Placement of Displays by Empirical Analysis of User Behavior. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.* ACM, 611–621.
Shafaq Mussadiq, Rehan Hafiz, and Muhammad Abdullah Jamal. 2017. Confined projection on selected sub-surface using a robust binary-coded pattern for pico-projectors. *Multimedia Systems* 23, 2 (2017), 207–222.
George Papandreou, Tyler Zhu, Nori Kanazawa, Alexander Toshev, Jonathan Tompson, Chris Bregler, and Kevin Murphy. 2017. Towards accurate multi-person pose estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 4903–4911.
Hyocheol Ro, Jung-Hyun Byun, Inhwan Kim, Yoon Jung Park, Kyuri Kim, and Tack-Don Han. 2019. Projection-Based Augmented Reality Robot Prototype with Human-Awareness. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 598–599.
Hyocheol Ro, Inhwan Kim, JungHyun Byun, Yoonsik Yang, Yoon Jung Park, Seungho Chae, and Tackdon Han. 2018. PAMI: Projection Augmented Meeting Interface for Video Conferencing. In *2018 ACM Multimedia Conference on Multimedia Conference.* ACM, 1274–1277.