# REALTIME PLANE DETECTION FOR PROJECTION AUGMENTED REALITY IN AN UNKNOWN ENVIRONMENT

Dongchul Kim<sup>1</sup>, Seungho Chae<sup>1</sup>, Jonghoon Seo<sup>2</sup>, Yoonsik Yang<sup>1</sup>, Tack-Don Han<sup>1</sup>

Department of Computer Science, Yonsei University, Seoul, Korea<sup>1</sup> Software Platform R&D Lab., LG Electronics Advanced Research Institute, Seoul, Korea<sup>2</sup> {Ddckim, Seungho.chae, Joonghoon.Seo, Yoonsikyang, Hantack}@msl.yonsei.ac.kr

#### ABSTRACT

We propose a realtime plane detection method for projectionbased Augmented Reality (AR) system in an unknown environment. While previous works usually designate space, the plane detection method automatically detects multiple planes based on the proposed constrained sampling strategy in RAndom SAmpleing Concensus (RANSAC). For each plane, an area for projection is selected for contents while considering occlusions by other objects. In addition, when the multiple planes are detected, the importance for contents is measured by the score functions based on the properties of planes such as size, color, and position. The proposed method can guide users to select plane for projection by visualizing the importances, or can automatically select a plane according to the users. We achieves a significant improvement in speed (about 260 times faster than the RANSAC) and high precision. These technique has become widely utilized in various AR applications such as AR game, and etc.

*Index Terms*— Signal processing for Smart space, Projection AR, Projector-camera System, Plane detection, Internet of Things (IoT)

# 1. INTRODUCTION

Steerable Augmented Reality (AR) provides information and enhances immersion of users by overlaying physical spaces with virtual contents [1]. To realize this concept, projectors have been widely used for augmenting virtual content. In addition, a camera is normally attached to the projectors in order to recognize objects and user. From this information, users can interact with the augmented contents. Thus, the projection AR system has been receiving increasing attention to realize AR and ubiquitous computing [2]. In this paper, we focus on projector-camera systems for realizing steerable AR, especially when the space information is an unknown.

Previously developed projector-camera systems can be divided into two types depending on the mobility of the projectors. Firstly, the location of projector is fixed because the space for augmentation is usually pre-designated. Using these fixed projectors, the original functions of specific objects are extended while overcoming their limitations. For example, Rekimoto *et al.* [3] utilizes two projectors and cameras. The two projectors and one camera are static, and the other camera is a pan-tilt camera. They allow interactions between users and the augmented contents within a designated space. They extend individual computers as environmental computers by using nearby planes. Jones *et al.* [4, 5] extend the projection space by augmenting the scenes on rear walls or background in order to enhance the gaming experience.

Secondly, projectors are oriented by utilizing a motorized platform that can be used for contents at any point in a room [2]. Consequently, objects in larger area can be handled field-of-view (FOV) with a project-camera system. For example, the Everywhere Displays [6] is a steerable projection display constructed with a pan-tilt mirror attached to the front of the projector. To correct projective distortion problems, visual marker patterns are projected and then the pattern is recognized with an RGB-camera. It has several applications as virtual content can be projected on walls and tables. Butz et al. [7] propose a book searching system using the ARToolkit [8] with a projector-camera system. Molyneaux et al. [9] detect and track moving objects using the SIFT descriptor [10] for determining projection surfaces, and display information on the surfaces of the objects. Recently, Wilson et al. [2] proposed the Beamatron system that consists of a pan-tilt system attached to the ceiling.

In this paper, a realtime plane detection method for projection-based AR system in an unknown environment with multiple planes. The main contributions of this paper can be summarized as follows. First, a realtime multiple plane detection method with constrained sampling, which is a main procedure in RANSAC, is proposed. Second, from the detected planes, a projection area is automatically selected within each plane by handling the occlusion due to real objects. Finally, the best plane (i.e., projection area) for projecting content is automatically selected while considering the size, ratio, and color of planes. Because all these procedures do not require any installation by the users, the proposed method can be utilized in any unknown environment.

The remainder of this paper is organized as follows. In

Section 2, we describe the proposed method in detail. Section 3 presents the experimental results and discussion. Finally, in Section 4, our conclusions are presented and we briefly discuss future work.

# 2. PROPOSED METHOD

Fig 1 represents the overall framework of the proposed system that mainly consists of four modules, namely, calibration, plane detection, projection area selection, and plane selection. The positions of the RGB-D camera, and projector are different, the data should be corrected to easily obtain correspondences. Thus, the relative 3D poses obtained from the RGB-D camera, and projector are computed in the calibration module. Then, planes are detected from 3D point clouds in realtime based on the proposed constrained sampling. Considering occlusions by other objects, projection area is determined for each plane. Finally, the best plane for projection is automatically selected in plane selection module.



Fig. 1. Overall framework of the proposed system.

## 2.1. Projector-camera System Calibration

Fig 2 shows the hardware setup of the proposed system. To set similar viewpoints between an RGB-D camera and a projector, the RGB-D camera is attached to the projector.



Fig. 2. Various sensors of the projector-camera system.

Our system utilizes various sensors, and the placement of all sensors in the 3D space is different. Thus, the 3D points in each device have different positions in each local coordinate system. Therefore, data from each sensor is necessary to be transformed into a common coordinate system. To address issue, calibrations between the sensors are performed in order to obtain intrinsic and extrinsic parameters of the hardwares. In our system, the relative position of the RGBD camera and projector is computed by using corner points on a checkerboard. Computed data can then be transformed into any local coordinate system because of the known extrinsic and intrinsic parameters.

#### 2.2. Multiple Plane Detection using Constrained RANSAC

For reliable plane detection in 3D map, we refine the depth map by using an edge preserving filter [11]. From the refined depth map, 3D point clouds of a scene can be obtained. The next step is to detect planes to determine the projection space from the 3D point clouds. First, we attempted to use the RANSAC; however, its computation complexity was very high in order to obtain sufficient results as it has high probability of selecting points on a different plane. Thus, we modify the sampling methods strategy of the RANSAC with two constraints based on the Gestalt grouping [12] as follows:

C-(1) Points in local areas have a high possibility to be on the same plane.

C-(2) Points within an object have similar color values.



**Fig. 3**. Modified sampling with two constraints: (a) Sparse sampling using a rasterized scanning method; (b) Local sampling based on two constraints.

Fig 3 shows an example of the proposed sampling method based on the two constraints. While the RANSAC randomly selects samples from the entire point cloud, the proposed sampling randomly select a single point and other points are selected based on the constraints. The proposed sampling consists of two steps. First, a reference point is randomly selected by using a rasterized scanning method. In Fig 3(a),  $p^{refer}$ , which is indicated by the red point, is selected as the reference point among the yellow points. Then, according to C-(1), two circular support regions are generated with radius r and r/2, as shown in Fig 3(b). Each support region is divided into  $360/\theta$  cells, and the mid-points of the arcs in all cells are the sample candidates for obtaining a plane with  $p^{refer}$ . As the minimum number of points for constructing a plane is three, the mid-points of the arcs are visited in a specific order until two points are selected among the sample candidates.

According to C-(2), a point  $p^s$  is selected where the color difference between  $p^{refer}$  and  $p^s$  is lower than a certain threshold value T. In addition, the points on the outer support region have a higher priority to be selected because they are farther from  $r^{refer}$  as compared to the points on the inner



Fig. 4. Perspective distortion correct process.

support region. The numbers in Fig 3(b) represent the search order for selecting the samples method. If two points are not selected, the positions to be checked are rotated by  $\theta/2$ , represented as green points. Finally, the points indicated by red circles are selected as samples of the reference point  $p^{refer}$ , shown in blue color.

From the selected samples, a plane parameter can be estimated. Then, the outlier and inlier points can be determined from the 3D point clouds based on the distance from the plane. We assume that a correct plane  $\pi_1$  is found when the number of inlier points is greater than threshold N. In order to detect multiple planes, the proposed sampling and plane parameter estimation are performed again after removing the inliers of  $P_1$  from the 3D point cloud. Finally, we can efficiently detect multiple planes  $\Pi = {\pi_1, \pi_2, ..., \pi_i}$  using the proposed sampling method.

#### 2.3. Projection Area Selection

When a projector-camera system moves, two problems are encountered because of the changes in pose. First, the augmented contents can be distorted or occluded by other objects that have different depths. In addition, when the contents are projected on the planes directly, perspective distortion occurs from the viewpoint of the user.

To solve these problems, we propose selection of the projection area within a plane while considering the occlusion problem as well as the perspective distortion. The perspective distortion is caused by the orientation difference between the detected planes ( $\Pi$ ) and the projector. Thus, to correct this orientation difference, we first place virtual cameras in front of  $\Pi$  (Fig 5(a)). Then, the point cloud of the plane is projected into the virtual camera. So, we can obtain an image of the front view (Fig 5(c)) from the depth map (Fig 5(b)).



Fig. 5. Perspective distortion.

From the image of the front view, we need to select the projection area. A rectangular area of maximum size is selected from the front view image and Fig 4 shows the overall process for finding the maximum size of the rectangle. The front view image contains many holes. Thus, a dilatation operation is performed in order to fill these holes (Fig 4(a)). In order to remove the outliers, we apply a connected component process, and determine the largest component as the inlier points and the remaining components as outlier points (Fig 4(b)). Then, in order to efficiently select the rectangle, we utilize an integral image to compute the area of the arbitrary rectangle quickly (Fig 4(e)). In addition, canny edge detection [13] is also used; it uses them as starting points to find the inner rectangles (shown as red points in Fig 4(d)). For each start point, the searching direction is determined while considering the object orientations to further improve the computation. Since the orientations of planes are not known, we use the principal components to obtain the orientations of the planes using principal component analysis (PCA). In Fig 4(d), the two blue lines denote the estimated orientations of the green plane. Then, the plane is divided into four parts based on the blue lines in order to determine the search directions of the red points.

However, the computation area using an integral image is valid when the two sides of the rectangle are parallel to the horizontal and vertical axes in the image coordinate system. To overcome the limitation, the rotation images of the maximum sized blob are generated (Fig 4(c)). Finally, the maximum rectangle is selected from all rotated images (Fig 4(f)).

#### 2.4. Proposed Projection Area Selection Method

From the results of Section 2.3, the projection area for each detected plane  $\pi$  is determined. However, in AR applications, multiple planes are usually detected, and the planes for projection should be selected for an unknown environment.

For this, we consider the properties of planes such as width/height ratio, color, and area in a 3D space. It is better if the ratio is closer to the projector ratio because the projected contents do not need to be expanded; moreover, it is better if the color is similar to white because white reflects the beams of a projector well. Finally, the area of the plane should be big to project contents for immersive interaction.

We define three different score functions to consider each

Scene #		RANSAC	Proposed		RANSAC	Proposed		RANSAC	Proposed
1	1	9025.1	25.4	1	90.0	91.2		1545.7	7.7
2		1017.3	126.7		92.5	93.2	ms	1785.7	7.6
3	l III	26888.2	118.1	%) 	87.0	87.6	le (	2622.5	25.9
4	on co	41656.0	123.6	sion (	92.7	90.6	g Tin	5975.1	25.8
5		30159.7	120.7		95.0	94.6		6638.3	33.5
6	ati	79696.0	35.9	eci	86.4	87.1	Sin	7070.0	7.3
7	iter	12566.2	175.4	P	90.2	89.2	ces	1073.6	36.0
Avg.	1 -	28715.5	103.7	1	90.5	90.5	Pro	3815.8	20.5

Table 1. Comparison of the proposed method with RANSAC in terms of iteration count, precision, and processing time.



**Fig. 6**. Results of the proposed method: (a)(c) projective distortion scene; (b)(d) corrected perspective distortion



Fig. 7. Processing time of the proposed method.

property, and the objective function is defined to select a plane from  $\Pi$ .  $S^a$ ,  $S^c$ , and  $S^r$  represent score values considering size, color and ratio of the projection area.

In Eq.(1),  $\pi^a$ ,  $\pi^c$ , and  $\pi^r$  are the area, mean color, and width/height ratio of a plane  $\pi$ . In addition,  $\Psi^r$  is the projection ratio of a projector. Larger score value represents a better plane for projection for all score functions.

$$S^{a}(\pi) = -\exp\left(-\frac{\pi^{a}}{2}\right) + 1,$$
  

$$S^{c}(\pi) = \exp\left(-\beta \frac{\sum_{c \in \{r,g,b\}, \pi \in \Pi} |\pi^{c} - 255|}{255 \times 3}\right),$$
  

$$S^{r}(\pi) = \max\left(-\gamma \left(\pi^{r} - \Psi^{r}\right)^{2} + 1, 0\right),$$
(1)

# 3. EXPERIMENTS

The experiments were conducted using a 2.40 GHz CPU, 8 GB memory. The image resolution was  $640 \times 480$ . To verify the proposed method, we compare the proposed method with the RANSAC method. Since no standard dataset is available that depth map and labels of planes at the same time, we manually generate ground truth data of 7 difference scenes to measure the quantitative precision.

Fig 6 and Fig 7 shows the results images and the processing time of the proposed method. The depth refinement requires about 10 ms. The processing time of the projection area determination and plane selection require approximately 5 ms. We can check that the processing speed of the proposed method fast in realtime. Fig 6 shows the result scene obtained through the proposed method. Also, we compared the proposed method with the RANSAC using two quantitative measures using Eq.(1) (See Table 1).

## 4. CONCLUSION

We propose a realtime plane detection and selection method for a projector-camera AR system in an unknown environment. The proposed method achieves significant improvement in terms of speed (about 260 times faster than the RANSAC) while retaining precision (about 90.5%) similar to that of RANSAC. In addition, when multiple planes are detected, the importances of projecting contents is measured by the proposed three score functions based on the properties of the planes such as ratio, color, and area. In addition, the steerable projection AR hardware prototype is implemented to offer a pervasive display environment for user spaces. This system can create a pervasive display environment on user spaces and can adapt in realtime. The proposed system has a simple structure and is easy to install, making it easily applicable in practical life. We believe that the proposed method can be widely utilized in various AR scenarios such as smart table, video conferencing, and immersive interaction.

# 5. ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. NRF-2015R1A2A1A10055673).

# 6. REFERENCES

- Ramesh Raskar, Greg Welch, and Henry Fuchs, "Spatially augmented reality," in *First IEEE Workshop on Augmented Reality*. Citeseer, 1998, pp. 11–20.
- [2] Andrew Wilson, Hrvoje Benko, Shahram Izadi, and Otmar Hilliges, "Steerable augmented reality with the beamatron," in *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 2012, pp. 413–422.
- [3] Jun Rekimoto and Masanori Saitoh, "Augmented surfaces: a spatially continuous work space for hybrid computing environments," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 1999, pp. 378–385.
- [4] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson, "Illumiroom: peripheral projected illusions for interactive experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 869–878.
- [5] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira, "Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units," in *Proceedings of the* 27th annual ACM symposium on User interface software and technology. ACM, 2014, pp. 637–644.
- [6] Claudio Pinhanez, "The everywhere displays projector: A device to create ubiquitous graphical interfaces," in *Ubicomp 2001: Ubiquitous Computing*. Springer, 2001, pp. 315–331.
- [7] Andreas Butz, Michael Schneider, and Mira Spassova, "Searchlight–a lightweight search function for pervasive environments," in *Pervasive Computing*, pp. 351–356. Springer, 2004.
- [8] Hirokazu Kato and Mark Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on. IEEE, 1999, pp. 85–94.
- [9] David Molyneaux, Hans Gellersen, Gerd Kortuem, and Bernt Schiele, *Cooperative augmentation of smart objects with projector-camera systems*, Springer, 2007.
- [10] David G Lowe, "Distinctive image features from scaleinvariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [11] Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother, and Margrit Gelautz, "Fast cost-volume

filtering for visual correspondence and beyond," in *Computer Vision and Pattern Recognition (CVPR)*, 2011 *IEEE Conference on*. IEEE, 2011, pp. 3017–3024.

- [12] Kuk-Jin Yoon and In-So Kweon, "Locally adaptive support-weight approach for visual correspondence search," in *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, vol. 2, pp. 924–931.
- [13] John Canny, "A computational approach to edge detection," 1986, number 6, pp. 679–698, IEEE.