AR sightseeing system with proximity detection and markerless AR

Hidetaka MIYAJIMA, NorikoYATA, Yoshitsugu MANABE
Graduate School of Advanced Integration Science
Chiba University
Chiba, Japan

***Abstract*—This paper proposes a sightseeing guidance system using a beacon and markerless AR which is robust with time. This system provides users with sightseeing guidance by AR without spoiling the scenery.**

Keywords— AR; beacon; sightseeing guidance system

I. INTRODUCTION

In recent years, augmented reality (AR) technology has been widely spread in various fields such as entertainment. AR is a technique which adds a virtual information generated by computer to the real world information perceived by human beings. This technology has been used as a new guidance method in sightseeing spots. For such the system, a location based AR using such as GPS information and a marker based AR are generally used. However, there are problems such as accuracy of position estimation according to equipment and spoiling the scenery by attaching markers in sightseeing spots. Therefore, this paper proposes a sightseeing guidance system more suitable for sightseeing spots by combining a beacon that can detect user's proximity information with high accuracy and markerless AR that does not require attaching markers.

II. SIGHTSEEING AR APPLICATION

It is considered effective to use markerless AR because the marker based AR which requires attaching markers in the sightseeing spots spoils the scenery. The markerless AR is a technique which matching arbitrary feature points with feature points in a camera image to extract a recognition target from image information and track it. However, in the case of the markerless AR, there arises a problem that the user can not know where and what to do because there is no marker. Therefore, by solving this problem a beacon is used to detect the user at the place where is requested to use the markerless AR. When the user enters this area, a message like "Please hold the smartphone over the building" is announced from the digital device. The user holds the device in accordance with the instruction. The system recognizes the target such as highlights of sightseeing spots, and superimposes a virtual information on the device.

At the same time, outdoor sightseeing spots have places hidden by crowds, places where shadows and plants change due to time influences as shown in Fig. 1. Matching cannot be performed correctly if feature points for markerless AR are detected from those parts. In this case the system cannot be recognized the target robustly. Therefore, this research finds invariant feature points regardless of obstacles or time. By using this invariant feature points, the markerless AR that can be stably displayed at all times is created. Fig. 2 shows an image of the system.

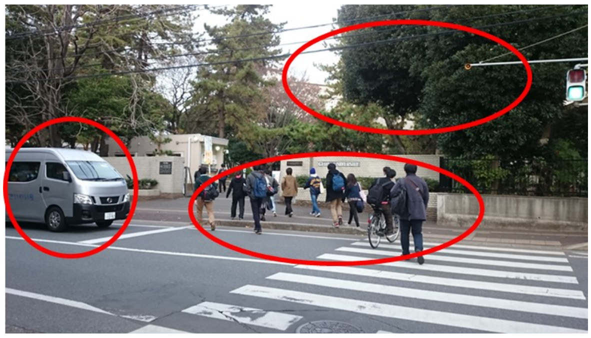

Fig.1. Influence of People and Time

Fig.2. Overview of System

III. DETECTING FEATURE POINTS

To recognize the target, feature points on the target are used. In this research, we use Accelerated-KAZE (AKAZE) feature [1]. In conventional SIFT feature [2] and SURF feature [3], since Gaussian filters are used, smoothing of important edge occurs. In contrast, since AKAZE feature uses a nonlinear diffusion filter, smoothing of the edge is suppressed, and feature extraction can be performed while maintaining the better features of the image. For this reason it has high

978-1-5386-2615-3/18/$31.00 ©2018 IEEE).

robustness against scale, rotation, and luminance change of the image compared with SIFT feature and SURF feature. Such characteristic of AKAZE allows to deal with appearance change of the captured objects due to the standing position and the height of the user within the detection area of the beacon.

IV. Invariant Feature Points

In this research, feature points detected at high frequency regardless of time in unchanged area such as buildings are defined as invariant feature points. In order to obtain invariant feature points, it is necessary to first extract a part which does not change with time from the image. Therefore, a mask of the part which does not change with time in the image is created. However, creating the mask manually for large numbers of images requires a great deal of effort. In this research, the mask is automatically created by using machine learning. First, as a teacher data for learning, as shown in Fig. 3, the mask can be created by the part of not change with the time in each image. In this research, training data is learned using FCN (Fully Convolutional Network) [4]. FCN is composed entirely of the Convolution layer and is a model used for pixel-wise segmentation. Based on the results of learning, mask generation is automated.

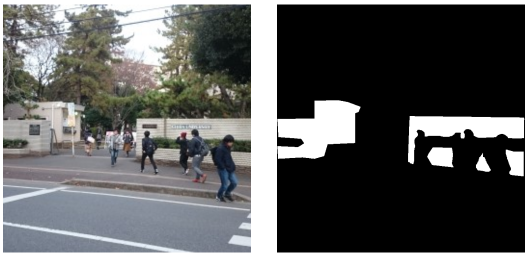

Fig.3. Mask Image

V. METHOD OF EXTRACTING INVARIANT FEATURE POINTS

The invariant feature points extraction method is shown below.

1. Prepare many images of the target taken in various time and season.
2. Generate the mask for the captured image.
3. Extract feature points from the mask area in each image.
4. Match feature points between images. When matched, record the number of matching times for each feature point.
5. Feature points with a large matched number of times are regarded as invariant feature points.

VI. EXPERIMENTAL RESULTS

The results of experiments on the accuracy of the beacon to be used, the generation of a mask image using learning, and the extraction of feature points from the mask area are shown. It also shows the results of calculating the internal parameters and the external parameters of the camera necessary for displaying the AR using invariant feature points.

A. Beacon

A beacon that constantly emits a radio wave of a certain frequency and receives the radio wave by the Bluetooth function of the device is used (Fig. 4). Based on the strength of the received radio wave, it is possible to measure how far the user is to the beacon. However, since the intensity of this radio wave varies with the influence of surrounding walls and objects, there are cases in which it is not possible to obtain an accurate distance. Therefore, it is necessary that the accuracy of the beacon with and without obstacles in the surroundings is verified. Table 1 and Table 2 show the results of comparing the distance calculated from the intensity of the radio wave and the actual distance. As shown in the tables, it was found that the errors and variance of the calculation distance become larger after passing 3 m in the space without obstacles and 2 m in the space with obstacles. In this research, the detection area of the user is set within the range of 2 m to 3 m according to the installation location of the beacon.

Fig.4. Beacon

TABLE I. WITHOUT OBSTACLES

Actual Distance (m)	**Calculated Distance (m)**	**Variance**
0.00	0.13	2.67
1.00	1.00	4.67
2.00	1.92	4.42
3.00	3.55	6.00
4.00	8.25	6.22
5.00	14.68	9.56

TABLE II. WITH OBSTACLES

Actual Distance (m)	**Calculated Distance (m)**	**Variance**
0.00	0.46	2.89
1.00	0.95	6.22
2.00	2.71	11.56
3.00	5.62	12.67
4.00	10.39	13.56
5.00	8.25	32.89

B. Generate mask image by learning method

Learning was prepared with 40 image sets, with 30 epochs and batch size 2. Fig. 5 shows the mask created by applying the learning result to the original image. As shown in the figure, the mask is made on the part avoiding people and plants.

Fig.5. Generated Mask

C. Result of extraction of invariant feature points

In this study, feature points are taken from mask area in 100 images, and feature points with frequent occurrence exceeding 60% are taken as invariant feature points. The extracted invariant feature points are displayed in Fig. 6. It was found that feature points were extracted from an area with height not be hidden when a person passed and not be shadowed.

Fig.6. Invariant Feature Points

D. Camera position and orientation estimation using invariant feature points

In displaying virtual information such as 3DCG, the internal parameters and the external parameters of the camera are necessary. The internal parameters were calculated by calibration using Zhang's method [5]. The external parameters are calculated using the PnP (Perspective n-Point) problem [6], [7]. By assuming invariant feature points to be in the Z plane in three-dimensional coordinates, it can be calculated by matching the three-dimensional point of the target with the two-dimensional image point detected by the device. Fig. 7 shows X (Blue), Y (Green), and Z (Red) axes of the world coordinate system viewed from the camera are drawn using the extracted invariant feature points and camera parameters. The image is captured the target almost from the front, but when showing the figure, the three axes are distorted. It is assumed that this occurs due to calculate using result including the incorrect matching.

Fig.7. World Coordinates Seen from the Camera

VII. CONCLUSION AND FUTURE WORKS

This paper proposed the sightseeing guidance system combining beacon and markerless AR. In order to realize the system, we verified the accuracy of the beacon, extracted invariant feature points using the AKAZE feature, and estimated the position and orientation of the camera using feature points. As a result, detectable range of beacons, invariant feature points, and camera position and orientation information were obtained.

In future, we will obtain more accurate camera position and orientation information by improving the accuracy of feature point matching. Furthermore, based on the camera's position and orientation information, CG is superimposed at an appropriate position. And we actually evaluate the system which combines beacon and markerless AR.

REFERENCES

[1] P.F.Alcantarilla, J.Nuevo, A.Bartoli, "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces", British Machine Vision Conference (BMVC), pp.13.1-13.11, 2013.

[2] David G Lowe, "Distinctive image features from scale-invariant keypoints", International journal of computer vision, vol. 60, no. 2, pp. 91-110, 2004.

[3] Bay Herbert, Ess Andreas, Tuytelaars Tinne SURF:speeded up robust features[J]. Computer vision and image understanding, 2008 110(3):346–359.

[4] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation", Proc. IEEE CVPR, pp. 3431-3440, 2015.

[5] Z. Zhang, "A flexible new technique for camera calibration", IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 11, pp. 1330-1334, 2000.

[6] Francesc Moreno-Noguer, Vincent Lepetit and Pascal Fua. "Accurate non-iterative O(n) solution to the PnP problem," Proceedings of IEEE 11th International Conference on Computer Vision, 14-21 Oct. 2007, pp: 1-8

[7] M.A. Fischler and R.C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM, 24, pp381-395, 1981.