2012 IEEE RO-MAN: The 21st IEEE International Symposium on
Robot and Human Interactive Communication.
September 9-13, 2012. Paris, France.

Furniture Layout AR Application using Floor Plans based on Planar Object Tracking

Taiki Fuji[1], Yasue Mitsukura[2], and Toshio Moriya[3]

Abstract— **In this paper, we propose a new approach of Augmented Reality (AR) system for the furniture layout based on a planar object tracking. The planar object tracking methods using natural features are effective methods to estimate the object's pose and position in the AR applications because we are able to use the natural images. However, most of the feature descriptors have a lot of matching procedure. Therefore, by using an efficient feature point descriptor which is very fast both to build and to match, we track the planar objects. Especially, we use floor plans as the planar objects, and then furniture CG models are overlaid on the floor plans. This is because the floor plans are presented in the selection of rooms for rent or buy. Therefore, this system helps borrowers and buyers to select some mansion or apartment rooms. In this system, we propose to use human whistle sounds and color rectangles recognition to operate the furniture layout. In order to show the effectiveness of our proposed system, we perform some planar object tracking experiments when we applied the proposed system to some floor plans.**

I. INTRODUCTION

Augmented Reality (AR) technologies enhance our perception and help us to see, hear, and feel our environments in new and enrich ways [1]. AR technologies are capable of applying various AR services (e.g. information services, video games, entertainments, support systems for field works, and so forth). In studies of AR, video see-through methods are major realization procedure because AR technologies are generally realized with a camera. Such methods overlay CG to the video frames of user's vision.

Recent years, we have seen advances in many enabling AR technologies. Furthermore, much research has been carried out on how AR can be used to enhance existing applications [2]. Many AR applications are developed by the visual markers which are some pictures or paintings in place of ARToolKit [3] markers. AR books [4] or board games [5] are the representative application of emerging AR. Therefore, the planar object tracking is needed for realizing the visual marker-based AR applications.

In addition to these AR application, furniture layout AR systems are expected to realize for consumer of apartment buildings. In studies of virtual furniture arrangements, various methods have been proposed. Germer and Schwarz present a procedural approach to generate furniture arrangements for large virtual indoor scenes [6]. This paper introduces an agent-based solution and demonstrates the flexibility and effectiveness of the agent approach. Merrell *et al.* [7] present a method for automated generation of building layouts for computer graphics applications. In this study, the architectural program is realized in a set of floor plans. The floor plans are used to construct complete three-dimensional buildings with internal structure. Yu *et al.* [8] present a system that automatically synthesizes indoor scenes realistically populated by a variety of furniture objects. This system realizes fully automatic furniture layouts using an optimization method with considering ergonomic factors. These CG based furniture arrangements are effective for the virtual reality environments or the time before the construction of the buildings. However, there are cases that the buyer of apartments or housings wants to lay out the furniture in real time and real environments. Therefore, applying AR technologies to the furniture layout systems is effective for the situations.

In related works, several marker-based alignment methods for the furniture layout AR applications. Billinghurst *et al.* [9] describe several interaction methods that can be used to provide a better user experience, including tangible user interaction, multi-modal input and mobile interaction. Moreover, some marker-less alignment methods for the furniture layout AR applications are presented. We previously presented a model-based alignment method for the furniture arrangement AR application in an indoor environment (e.g. some room) [10]. Koller *et al.* [11] present a landmarks-based tracking method in an indoor environment. These alignment methods assume that the furniture layout AR applications are used in some rooms of condominium buildings. However, purchasers see several floor plans with the persons in charge before purchasers visit the rooms. Therefore, a floor plan-based AR system is effective for the selecting the rooms which the purchasers are going to visit.

Generally, apartment sellers show some floor plans to buyers. However, most of buyers are hard to imagine how the rooms are. Therefore, the floor plans-based AR system helps for narrowing down the number of rooms which are selected if you come in virtually any image easily. Moreover, by placing the furniture CG models to the floor plan papers, some information of room space and surroundings are provided to the buyers.

For building the AR application, we need to track some

[1]T. Fuji is with Graduate School of Bio-Applications and System Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan, and a visiting research student of School of Integrated Design Engineering, Keio University, Kanagawa 223-8522, Japan. fuji@mitsu.sd.keio.ac.jp

[2]Y. Mitsukura is with Department of System Design Engineering, Keio University 3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa 223-8522, Japan. mitsukura@sd.keio.ac.jp

[3]T. Moriya is with Advanced Research Laboratory, Hitachi, Ltd. 2520, Akanuma, Hatoyama, Saitama 350-0395, Japan. toshio.moriya.tm@hitachi.com

floor plans which are printed in some papers. Many marker-less, or natural feature based methods are suggested in the AR. By using the planar object tracking, we track the target for the AR application.

In this paper, we propose the furniture layout AR application based on the visual marker-based alignment. We propose to use binary strings as an efficient feature point descriptor called BRIEF [12]. BRIEF is very fast both to build and to match. BRIEF descriptor is not designed to be rotationally invariant. However, it tolerates small amounts of rotation. Moreover, we also propose the AR interface for operating the AR application.

The major contributions of this paper are as follows:

- Applying furniture layout AR system to floor plans as visual markers
- High speed detection and tracking system without using GPU
- Interactive operation system of the furniture layout AR application.

II. VISION-BASED TRACKERS FOR AR APPLICATIONS

To combine the CG models geometrically in the right place of real space, alignments between real space and virtual space are needed. Many vision-based trackers for AR applications have been suggested, and can be divided into two types, i.e. marker-based methods and marker-less methods.

A. Marker-based Methods

There are a lot of marker methods as the conventional methods typified by ARToolKit [3] and ARTag [13] in vision-based alignments. ARToolKit is one of the early stage trackers widely used in AR applications. It relies on intensity thresholding to detect the markers, and template matching to recognize them.

B. Marker-less and Natural Features-based Tracking Methods

Marker-less or natural features-based methods without using the ARToolKit markers are needed. In these methods, natural images or pictures are used in place of ARToolKit type markers. Many marker-less, or natural features-based methods have been proposed. Local descriptors such as SIFT[14] and SURF[15] have been used but do not show enough performance for real-time AR applications due to their heavy computational costs. Randomized Trees (RT) [16] and Feans[17] are faster than the low speed of the SIFT algorithm. However, these methods require a training time of more than one minute per target. Therefore, we use BRIEF descriptor [12] for detecting the planar object target.

III. PROPOSED SYSTEM

Our proposed system is based on a natural image marker-based alignment using some floor plans. We use the BRIEF descriptor [12] for detecting the planar objects. BRIEF descriptor is high-speed descriptor as compared to SIFT and SURF descriptors. Then we track the feature points using optical flow for high speed tracking. In our proposed AR system, we use two methods for operating the application. One is human whistle sound and the other is color recognition.

A. Detection and Tracking Methods

1) Detection Method: In this system, we use BRIEF descriptor for detecting the planar object target. BRIEF descriptor [12] simply creates a bit vector out of the test responses, which we compute after having smoothed the image patch. Test τ on patch $\boldsymbol{p}$ of size $S \times S$ as

$$\tau(\boldsymbol{p};\boldsymbol{x},\boldsymbol{y}) := \begin{cases} 1 & if\ \boldsymbol{p}(\boldsymbol{x}) < \boldsymbol{p}(\boldsymbol{y}) \\ 0 & otherwise \end{cases} \tag{1}$$

where $\boldsymbol{p}(\boldsymbol{x})$ is the pixel intensity in a smoothed version of $\boldsymbol{p}$ at $\boldsymbol{x} = (u,v)^T$. Choosing a set of n_d $(\boldsymbol{x},\boldsymbol{y})$-location pairs uniquely defines a set of binary tests. BRIEF descriptor is taken to be the n_d-dimensional binary string

$$f_{n_d}(\boldsymbol{p}) := \sum_{1 \le i \le n_d} 2^{i-1}\tau(\boldsymbol{p};\boldsymbol{x}_i,\boldsymbol{y}_i). \tag{2}$$

When creating the descriptor, the only choices that have to be made are those of the kernels used to smooth the patches before intensity differencing and the spatial layout of the ($\boldsymbol{x}$, $\boldsymbol{y}$)-pairs. Eq. (1) takes only the information at single pixels into account, and it is therefore noise-sensitive. By Gaussian smoothing of the patch, this sensitivity can be reduced, thus increasing the stability and repeatability of the descriptor. According to [12], Gaussian kernel ranging is from 0 to 3. In practice, it uses a value of 2. Moreover, the smoothing kernel size of being necessary and sufficient is 9×9 pixels. Generating a length n_d bit vector leaves many options for selecting the n_d test locations ($\boldsymbol{x}_i$, $\boldsymbol{y}_i$) of Eq. (2) in a patch of size $S \times S$. Assuming the origin of the patch coordinate system to be located at the patch center, they can be described as follows. ($\boldsymbol{X}$, $\boldsymbol{Y}$) $\sim$ i.i.d. Uniform($-\frac{S}{2}$, $\frac{S}{2}$): The ($\boldsymbol{x}_i$, $\boldsymbol{y}_i$) locations are evenly distributed over the patch and tests can lie close to the patch border. Matching a number of points between two images typically involves three steps:

1) detecting the feature points,
2) computing the description vectors,
3) matching, which means finding the nearest neighbor in descriptor space.

In the case of BRIEF, any fast detector such as CenSurE[18] or FAST[19] can be used. In this paper, we use the FAST detector.

2) Tracking Method: To track the natural feature points, we use optical flow method. The optical flow of each image pixel as the distribution of apparent velocity of moving brightness patterns in an image. The flow of constant brightness profile can be described by constant velocity vector $\boldsymbol{v} = (v_x, v_y)^T$ as shown in Eq. (3),(4).

$$\begin{aligned} \boldsymbol{I}(x,y,t) &= \boldsymbol{I}(x+\delta x, y+\delta y, t+\delta t) \\ &= \boldsymbol{I}(x+v_x \cdot \delta t, y+v_y \cdot \delta t, t+\delta t) \end{aligned} \tag{3}$$

$$\frac{\partial \boldsymbol{I}(x,y,t)}{\partial t} = 0 \quad \Rightarrow \quad \frac{\partial \boldsymbol{I}}{\partial x} \cdot v_x + \frac{\partial \boldsymbol{I}}{\partial y} \cdot v_y + \frac{\partial \boldsymbol{I}}{\partial t} = 0 \tag{4}$$

The optical flow methods try to calculate the motion between two image frames which are taken at times t and $t + \delta t$ at every position. In particular, the Lucas-Kanade method [20] assumes that the displacement of the image contents between two image frames is small and approximately constant within a neighborhood of the point under consideration. By using the optical flow method, we can track the natural feature points in high speed.

3) Homography Estimation: By using the tracked feature points, we calculate the homography for estimating the target pose and position. A 3×3 homography matrix $\boldsymbol{H}$ between the input image and each reference image is calculated as follows:

$$c\begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} \sim \boldsymbol{H}\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{5}$$

where c is any non-zero constant. Then we use RANSAC[21] to robustly compute the homography $\boldsymbol{H}$. The method RANSAC can handle practically any ratio of outliers.

By using the homography and internal parameters of the camera, the camera rotation and translation can be easily calculated.

B. To Operate the AR Application

1) Operation using Human Whistle Sound: In the AR, real-time processing is demanded. Therefore, we would like to avoid intensive processing because AR system requires the CPU power and memory for other processing. The human whistle sounds are characteristic waveform and manageable as the feature amount [22]. The frequency spectrum of human whistle sounds is almost single and strong. Therefore, the human whistle sound interface works even if we only use a threshold processing in power spectrum. We show a wave form and frequency spectrum of the human whistle sounds using sound imaging application shown in Fig. 1.

The whistle sound has a nearly sinusoidal waveform. Therefore, the whistle has the strong single spectrum in the frequency domain. For this reason, the whistle sound is easily detected and recognized without using the complex speech analysis and implementations. In the proposed system, the two distinct pitch whistles (high/low) and three kinds of durations (short/middle/long) are used for operation cues (cf. TABLE I). In the durations, users sound the human whistle sounds with seeing the length barometer on the right bottom of the display. We simply process the human whistle sounds detection and pattern recognition. In the human whistle sound detection part, we calculate the power spectrum using fast fourier transform (FFT) by 0.1 [sec.]. The peak in the frequency domain of human whistle sound is intensive. Therefore, we easily obtain the frequency of the peak. Then, we judge the pitch and length of the sounds. The power spectrum peak of observed human whistle sound is judged with a simple algorithm. We find the max peak in the frequency domain from 500 Hz to 3000 Hz. Secondly, the power spectrum of the peak is compared with a threshold. The threshold is determined on an empirical basis. Then, the pitch of the peak is judged (i.e. High, Low, and None). Moreover, the length of the peak is counted when the whistle is sounding. If the judge result of the pitch is the human whistle sounds, we determine the sound pattern using the pitch and length.

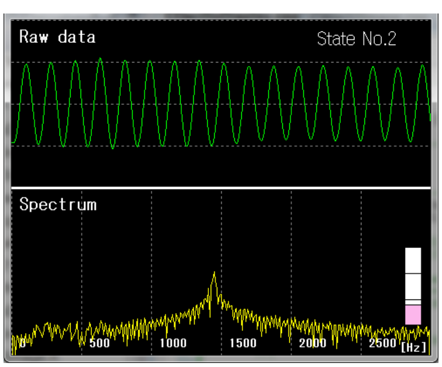

Fig. 1. Whistle sound application.

TABLE I
WHISTLE PATTERN.

Pitch \ Time Length	0.3-1.0(s)	1.0-2.0(s)	2.0- (s)
High (1500-2000Hz)	I	III	V
Low (1000-1500Hz)	II	IV	VI

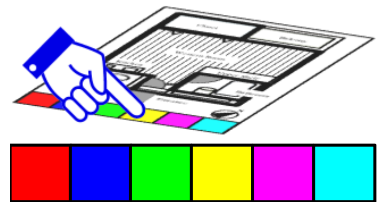

Fig. 2. Operation using color rectangles recognition.

2) Operation using Color Rectangles Recognition as Buttons: In color recognition, we also conduct a simple processing. Fig. 2 shows the rectangles of 6 colors (Red, Green, Blue, Yellow, Magenta, Cyan) at the bottom of the floor plans. The number of the operation patterns is up to factorial of 6, but we use a few patterns in this paper.

By concealing the color rectangles, the operation patterns are recognized. If the planar object targets are detected and tracked, these rectangle positions are automatically estimated. Therefore, we only estimate the color patterns which are able to be recognized.

IV. EXPERIMENTS USING FURNITURE LAYOUT APPLICATION

In this section, we describe our proposed furniture layout application using floor plan based alignment. A furniture layout application on some floor plans is an AR application which is used with a camera-equipped head mounted display (HMD). HMD is a typical device for AR applications. We

Fig. 3. Head Mounted Display (HMD) equipped with 2 cameras.

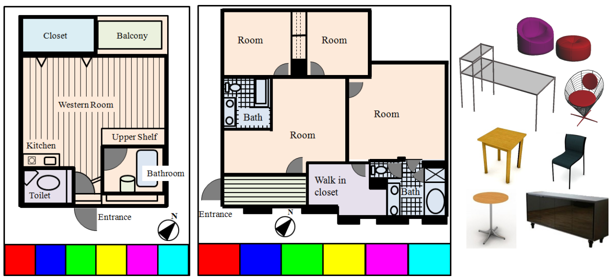

Fig. 4. Examples of the floor plans and furniture CG models.

use the video see-through type HMD (Wrap920AR from Vuzix corporation, cf. Fig. 3). Moreover, we assume that this application is used for imaging the room at the purchase of someone's home by the furniture layout.

A. Environmental Setting

We conduct experiments using the AR application which implemented the proposed method on a laptop PC with an Intel Core i7-2620M 2.7 GHz processor and 8.00 GB RAM running Windows 7. The resolution of the HMD's camera having 30 fps is set as 640×480 [pixels]. We use only the left eye's camera.

B. Detection and Tracking of the Planar Objects

Firstly, we show examples of the floor plans and furniture CG models in Fig. 4. A floor plan image includes a lot of feature points, for example, edges and corner points. Therefore, it is easy to detect and track the planar object tracking. Moreover, the targets of the floor plans were changed by mouse clicks of the top left point and the bottom right point. Furthermore, we compare the detection time of $n_d(\boldsymbol{x}, \boldsymbol{y})$-location pairs, that is, bit length. As previously described in Sec. III-A, the location pairs are randomly selected using uniform random numbers. In addition to the BRIEF algorithm, 18 steps rotation (by 20 degrees) and 3 steps scaling (0.5, 1.0, 1.5) are conducted. The detection rates are increased by changes in the templates. We show the detection time according to the length of the BRIEF descriptor was shown in TABLE II. As a result, we decide to use the 64 bits length considering the detection speed. Next, we show the detection results using BRIEF descriptor in Fig. 5. We can change easily the floor plans because the BRIEF descriptor is very fast both to build and to match. Moreover, we conducted the detection experiments using 64 bits and 256 bits lengths. Fig. 5 (a)-(d) show detection results using 64 bits length using the floor plans which is the left image of Fig. 4, and Fig. 5 (e)-(h) show detection results using 256 bits length using the floor plans which is the right image of Fig. 4. As a result, the detection performance using 64 bits length exponentially reduced at more than about 15 degrees because BRIEF is not designed to be rotationally invariant. Thus, in the detection phase, we needed to face the front of the floor plans. However, it was possible to detect the targets in near 90 degrees rotation when we used the 256 bits length.

TABLE II
DETECTION TIME ACCORDING TO THE LENGTH OF THE BRIEF DESCRIPTOR IN FRONT OF THE FLOOR PLANS.

n_d length	32	64	128	256
Frame rate (ave.) [fps]	more than 28	22	12	8

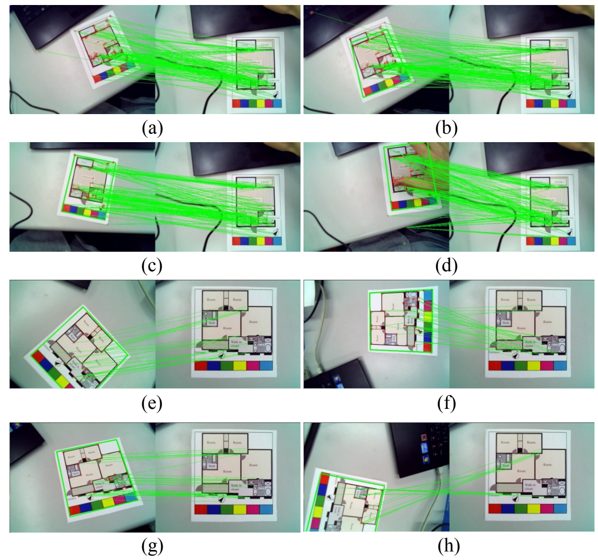

Fig. 5. Detection results using BRIEF Descriptor.

Our implementation ran at about 25 fps without using GPU when we used the 64 bits length.

C. Human Whistle Detection and Color Rectangles Recognition

In the manipulation system, we use the human whistle sounds and color rectangles recognition. We realized a number of interactive operations using color recognition which has a much lower cost in image processing. We used the color recognition to do translation, rotation and change of the furniture CG models. By using the human whistle sounds, we sensuously decided the degree of the translations and rotations in hands-free. Moreover, we were easy to change the furniture CG models.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed the furniture layout AR application based on the visual maker-base tracking methods. The proposed system made the interactive operation be possible

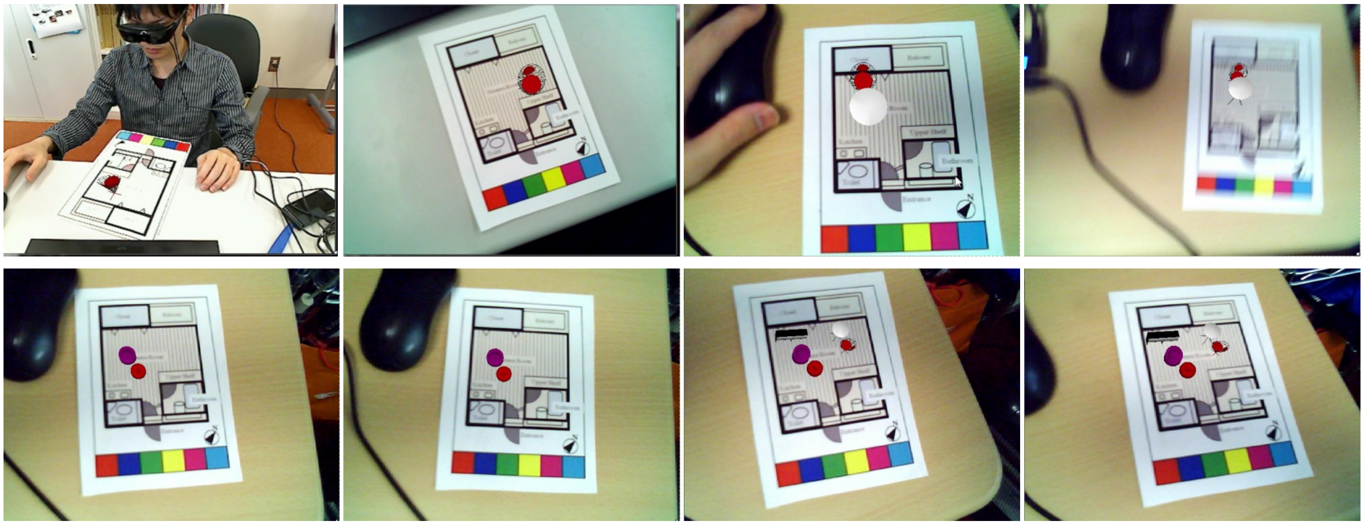

Fig. 6. Results of furniture layout AR application.

for the user using the camera-equipped HMD device. We demonstrated our approach on a furniture layout AR application using some floor plans. This AR application did not need GPU, and realized more than 20 fps in the detection. By using some floor plans and calculating its position and pose, we conducted the furniture layout on the floor plans. For operating the system, the combinations of the 6 colors recognition in complex with human whistle sounds were used. We think that our proposed AR system including the operation methods have a possibility to be applied to other AR applications (e.g. video games, AR books, and Supporting AR system in construction fields).

In future works, we need to consider the feedback of user's evaluation as a human interface. In AR studies, not only the estimation of the camera pose and position, the recognition of the real environments, and the AR interfaces are necessary for realizing the AR, but also the user's evaluation are very important issue. Moreover, we will use this system to watch over for avoiding dangerous layouts of furniture, considering the space between the furniture and rooms.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Japan Society for the Promotion of Science Fellows.

REFERENCES

[1] D. W. F. van Krevelen and R. Poelman, A Survey of Augmented Reality Technologies, Applications and Limitations, The International Journal of Virtual Reality, Vol. 9, No. 2, pp. 1–20, Jun. 2010.

[2] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht, An augmented reality museum guide, Proc. of Int. Symp. on Mixed and Augmented Reality (ISMAR '08), pp. 103–106, Sep. 2008.

[3] H. Kato and H. Billinghurst, Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, Proc. of IEEE/ACM Int. Workshop on Augmented Reality (IWAR '99), pp. 85–94, Oct. 1999.

[4] K. Kim, V. Lepetiti, and W. Woo, Scalable Real-time Planar Targets Tracking for Digilog Books, The Visual Computer: International Journal of Computer Graphics, Vol. 26, pp. 1145–1154, Jun. 2010.

[5] E. Molla and V. Lepetit, Augmented Reality for Board Games, Proc. of Int. Symp. on Mixed and Augmented Reality (ISMAR '10), pp. 1–8, Oct. 2010.

[6] T. Germer and M. Schwarz, Procedural Arrangement of Furniture for Real-Time Walkthroughs, Computer Graphics Forum, Vol. 28, No. 8, pp. 2068–2078, Jun. 2009.

[7] P. Merrell, E. Schkufza, and V. Koltun, Computer-generated residential building layouts, ACM Trans. on Graphics. Vol. 29, No. 6, Article 181, Dec. 2010.

[8] L. Yu, S. Yeung, C. Tang, D. Terzopoulos, T. F. Chan, and S. J. Osher, Make it Home: Automatic Optimization of Furniture Arrangement, ACM Trans. on Graphics, Vol. 30, No. 4, Article 86, Jul. 2011.

[9] M. Billinghurst, H. Kato, and S. Myojin, Advanced Interaction Techniques for Augmented Reality Applications, Proc. of Virtual and Mixed Reality (VMR '09), pp 13–22, Jul. 2009.

[10] T. Fuji, Y. Mitsukura, and T. Moriya, A Proposal of Model-based Alignment using Swarm Intelligence and Condensation, Proc. of Int. Symp. on Intelligent Signal Processing and Communication Systems (ISPACS '11), PID 222, Dec. 2011

[11] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, Real-time vision-based camera tracking for augmented reality applications, Proc. of ACM Symp. on Virtual Reality, Software and Technology (VRST '97), pp. 87-94, Sep. 1997.

[12] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzinski, C. Strecha, and P. Fua, BRIEF: Computing a Local Binary Descriptor Very Fast, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 34, No. 7, pp. 1281-1298, Jul. 2012.

[13] M. Fiala, ARTag, a fiducial marker system using digital techniques, Conf. on Computer Vision and Pattern Recognition (CVPR '05), Vol. 2, pp. 590–596, Jun. 2005.

[14] D. G. Lowe, Distinctive image features from scale-invariant key points, Journal of Computer Vision, Vol. 60, No. 2, pp. 91–110, Nov. 2004.

[15] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, SURF: Speeded Up Robust Features, Journal of Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346–359, Jun. 2008.

[16] V. Lepetit, P. Lagger, and P. Fua, Randomized trees for real-time keypoint recognition, Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR '05), vol. 2, pp. 775–781, Jun. 2005.

[17] M. Ozuysal, M. Calonder, P. Fua, and V. Lepetit, Fast keypoint recognition using random ferns, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 32, No. 3, pp. 448–461, Mar. 2010.

[18] M. Agrawal, K. Konolige, and M. Blas, Censure: Center Surround Extremas for Realtime Feature Detection and Matching, Proc. of European Conference on Computer Vision (ECCV '08), Vol. 5305, pp. 102–115, Oct. 2008.

[19] E. Rosten and T. Drummond, Machine Learning for High-Speed Corner Detection, Proc. European Conference on Computer Vision (ECCV '06), Vol. 1, pp. 430-443, May, 2006.

[20] B. D. Lucas and T. Kanade, An Iterative Image Registration Technique With an Application in Stereo Vision, Seventh Int. Joint Conf. on Artificial Intelligence, pp. 674-679, Aug. 1981.
[21] M. Fischler and R. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communcations of the ACM, Vol. 24, No. 6, pp. 381–395, Apr. 1981.
[22] T. Fuji, Y. Mitsukura, and N. Hamada, A Facile Interface for AR system using Human Whistle Sound, Proc. of The 2011 IEICE International Workshop on Smart Info-Media Systems in Asia (SISA '11), pp. 106-109, Nov. 2011.