Outdoor Augmented Reality using Optical see-through HMD System for Visualizing Building Information

Munetoshi Miyake[1], Tomohiro Fukuda[2], Nobuyohsi Yabuki[3], Ali Motamedi [4] and Takashi Michikawa[5]

1) Master Course Student, Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan. Email: miyake@it.see.eng.osaka-u.ac.jp
2) Ph.D., Assoc. Prof., Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan. Email: fukuda@see.eng.osaka-u.ac.jp
3) Ph.D., Prof., Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan. Email: yabuki@see.eng.osaka-u.ac.jp
4) Ph.D., JPSP Research Fellow, Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering,
Osaka University, Suita, Osaka, Japan. Email: motamedi@see.eng.osaka-u.ac.jp
5) Dr. Eng., Assistant. Prof., Center for Environmental Innovation Design for Sustainability, Osaka University, Osaka, Japan. Email: michikawa@ceids.osaka-u.ac.jp

Abstract:

With the development of Information Technologies (IT), a need for new methods to visualize digital information has appeared in architectural and urban field. Augmented Reality (AR) technology, that overlays computer graphics on real world scenes, can be used to satisfy the need. AR-systems can use a wide variety of displays with different characteristics and limitations. For example, most common AR mobile applications, that use devices such as smartphones, occupy both hands of the operator. On the other hand, Head Mounted Displays (HMD) allow more freedom of movement to the user. In this paper, an outdoor AR system for visualizing building information is proposed and implemented. The proposed system uses optical see-through HMD to facilitate visualization, conception and communication about building information by augmenting virtual models on real scenes while giving the users free movement capability. To realize the system, a marker-less AR, which was based on local feature image registration and Structure from Motion (SfM) technologies, was developed. In our proposed method, in the preprocessing step, using a number of photographs from several viewpoints, a point cloud representation of a building is reconstructed by SfM. Then, in the main process, the local features of the live video image of HMD are extracted in real time and are associated with photographs for SfM using image recognition algorithms (such as SURF). Finally, a label which has information for assessing and rating the environmental performance are rendered in the optical see-thorough HMD.

Keywords: Optical see-through Head Mounted Display, outdoor Augmented Reality (AR), local feature registration

1. INTRODUCTION

In recent years, information technology has been used throughout the whole life cycle of a facility. The life cycle includes planning, design, construction, operation and maintenance, management, and refurbishment. Information technology, such as Building Information Modeling (BIM), has been mainly used by designers and constructors, yet, it has begun to involve wider range of stakeholders. However, inefficient use of information tools sometimes causes misunderstanding between professionals and people who are unfamiliar with the tools. Hence, there is a need to develop new methods to help stakeholders understand building information, communicate with each other, and share ideas more efficiently.

The use of Augmented Reality (AR), which can visualize the design of a building, has started to be investigated in architectural and urban field (for example, Reitmayer et al., 2006; Karlekar et al., 2010; Schubert et al., 2015). AR overlays the images or 3D virtual objects with computer generated data on real world scenes through video or photographic displays. AR can help visualizing extra digital information of full scale building projects (Watanabe, 2011) and can give annotation on an actual building (Wither et al., 2009). AR is similar to Virtual Reality (VR) technology. VR needs to create not only target objects but also surrounding objects such as buildings and terrains. AR can get these surrounding objects from live video images, and only needs to add target objects, which makes it more time and cost saving. For a VR application, input devices are required to operate VR virtual camera by fly-through and walk-through. In contrast, in AR scene, this operation can be realized by a camera on a device, which is more intuitive for the user.

In this study, AR system is proposed to be used for an outdoor environment specifically in the operation and maintenance phase of a building's lifecycle. The proposed system shows users additional information about the condition of a building using augmented 2D labels. Our system enables non-professionals to check building information by integrating virtual objects in real-world scenes in the operation and maintenance stage. The proposed system can be used outdoors by users who are not expert in working with IT tools.

To realize the proposed AR system, one of the most significant problems is the display selection. Although high speed calculation can be easily executed in stationary computing devices and be displayed in desktop monitors, these monitors can only be used at their location and do not provide any mobility to the user. Held hand displays such as smartphone with camera can be easily accessed and used. Although they provide great mobility, the viewpoint from these devices is from their camera which is different from the actual viewpoint of the user and stereopsis is difficult to achieve when using these devices. In our system, Head Mounted Display (HMD) is utilized which offers high mobility and can provide a viewpoint from near human eyes.

There are two main types of HMDs that are compared in details in Rolland et al. (2000). One is video see-through HMD in which the real world scene is captured with video cameras mounted on the head gear or helmet and the computer generated images are electronically combined with the video representation of the real world. Although, registration between real world and virtual objects is relatively accurate, there are limitation of resolution and system delay on the video stream. Another option is optical see-through HMD in which the real world is seen through half transparent mirrors placed in front of the user's eyes similar to glasses. Mirrors are also used to reflect the computer generated images into the user's eyes and combine the real and virtual world views. In comparison to video see-through HMD, users can see real world through half transparent mirrors and there is no delay with real word vision.

Another big challenge to realize outdoor AR is achieving high accuracy registration. The accuracy mainly depends on geometric registration between live video and virtual models. To get a high accuracy, relationships between the position of users, posture of users and a position of drawn virtual objects should be precisely acquired by the system. The reference point of geometric registration is defined by sensor's position in sensor-based method (with physical sensors such as GPS, gyro sensor, acceleration sensor) or position of an artificial marker in marker-based methods (one of the vision based methods) (Yabuki et al., 2011). Using sensor-based methods, the registration accuracy depends on the accuracy of sensors which is not always high. For example, in an urban area, it is known that the GPS accuracy is not stable due to signal occlusion by high-rise buildings. In artificial marker-based methods, it is needed to attach a large marker the building to position the user. This method is straightforward, yet the camera for AR needs to capture the marker. Hence, there are limitations for user's movability range and challenges regarding installation of large markers.

Therefore, in this research a marker-less AR system, which does not need artificial markers or sensors, has been developed. Basic method of maker-less AR system relies on natural features of the images acquired from the live video. The position and posture of 3DCG model are decided according to these natural features. However, it is essential that natural features, which are used for geometric registration, are captured by the AR camera (such as HMD camera). Using point clouds for natural features is an option (such as the research by Yabuki et al. (2012)), however, the system requires special equipments such as 3D laser scanner to get point clouds. Structure from Motion (SfM) technique which reconstruct 3D shapes from photographs from plural view point is another method to create point cloud data (Dellaert et al., 2000), which is relatively easy and low cost.

In this study, a marker-less AR system by using optical see-through HMD with local feature-based image registration technology and SfM technology is developed. It offers less limitations, efforts and expenses for outdoor AR while providing the free movement. The developed system has extended the system developed by Sato et al. (2016) which applied to renovation design project.

2. SYSTEM OVERVIEW

An overview of the proposed system is shown in Figure 1. As shown in the Figure, the system has preprocessing and main processing: In the preprocessing step, a 3D model of an existing environment is digitalized by SfM using photographs taken from a number of viewpoints. In the SfM processing, the position and posture data of all photographs are saved to a database. These data are called camera parameters. Secondly, locations of designed virtual objects are relatively defined to the 3D model reconstructed by SfM. These defined coordinates and the 3D virtual objects are also stored in the database. Finally, key points and features from all the photographs used in SfM are extracted and saved as a text file. Figure 2 shows the data items used in our proposed method.

In the main processing, the data created in the preprocessing are imported to the system. Then, live video images are compared with photographs stored in the database through a matching process using local features. Once live video image matches photographs used in the SfM, the system uses camera parameters of matched photographs and estimates the position and posture and defines drawing points of the virtual object on the display. Details about each step are described in the following Subsections.

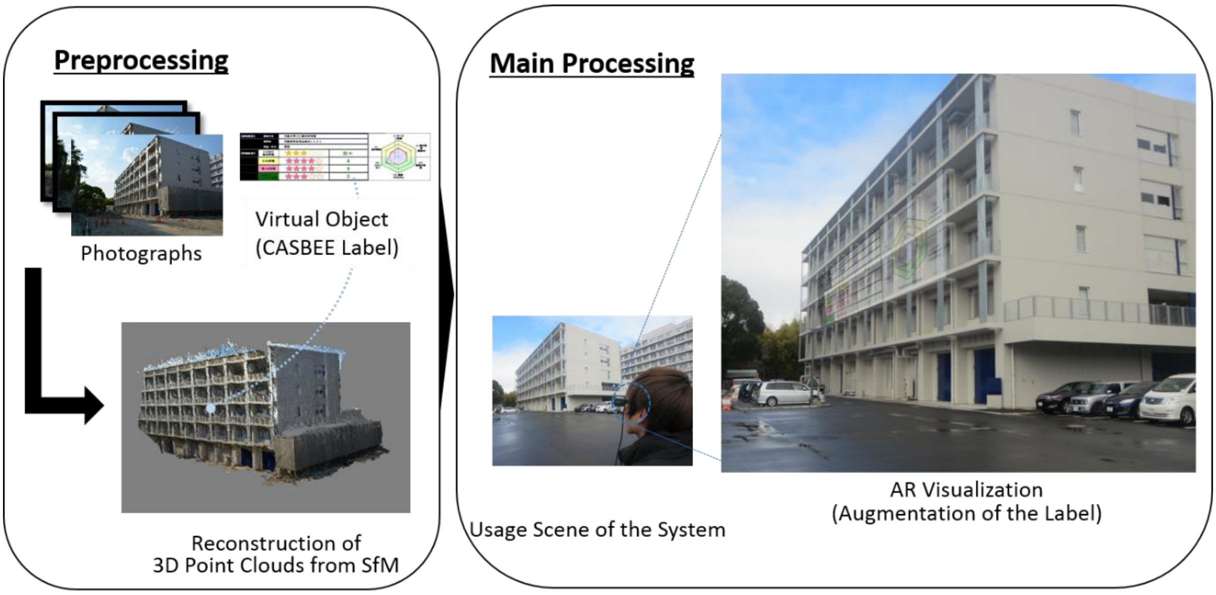

Figure 1. Overview of the system

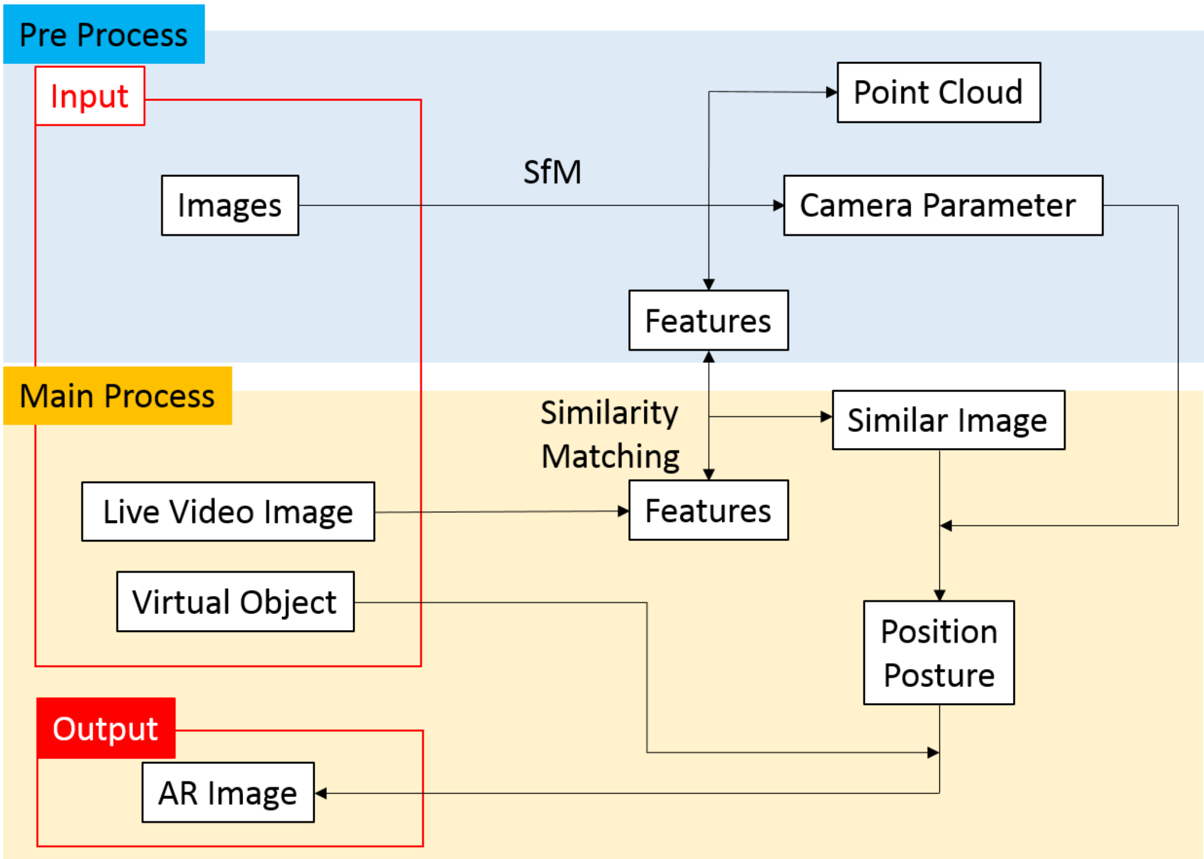

Figure 2. Marker-less AR data flow

2.1 Camera Parameter

Developed system needs camera parameters which are used to define the position of augmented virtual object on the 2D display. This translation is done using Equation (1) (Zhang, 2000). In this equation, s denotes the scaling factor, u and v are the position of object drawing on the screen coordinates. The value X, Y and Z were defined as the position of the object on the world's coordinates. The value f denotes the focal length, and c defines the coordinates of the center of the image. The value A is called internal parameters of a camera. The value R is defined as rotation matrix and the value T defined as translation matrix of a camera. The value $(R|T)$ is called external parameters of a camera.

$$s\overrightarrow{P_{screen}} = A(R|T)\overrightarrow{P_{world}} \qquad A = \begin{pmatrix} f_x & 1 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$A = \begin{pmatrix} f_x & 1 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \qquad R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \tag{1}$$

$$\overrightarrow{P_{world}} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

Internal parameters can be calculated through camera calibration and are specific to each camera. However, external parameters are specific to each photograph's viewpoints. Therefore, the system needs to re-calculate external parameters if a view point is changed. In this study, SfM is used to calculate external parameters of viewpoints for each photograph.

2.2 Reconstruction

The developed system needs a 3D model of the building. This 3D model is conventionally created using 3D modeling software (such as 3D CAD) that is usually a high cost activity. By using SfM method, the high cost issue can be solved. SfM is a technique to estimate three dimensional objects from two dimensional image sequences such as pictures. It also extracts camera parameter data which includes position and posture information of a camera. Using this technique, 3D model of the actual building and its environment can be easily constructed from photographs. To get a robust reconstructed 3D model, multiple pictures from various locations and view angles should be taken to cover the whole building surface. Data of reconstructed 3D virtual object and camera parameters for each photograph are stored in the database. The relative coordinates of new 3D virtual objects to the coordinates of 3D models reconstructed by SfM are defined and stored in the database. Key points and features of each photograph used for SfM are extracted and saved in a text file. In this study, an advanced local features detector descriptor namely Speed-Up Robust Features (SURF) is used. SURF is an advanced version of Scale-Invariant Feature Transform (SIFT) that is used for feature detection and image descriptor identification.

2.3 Distortion Correction

In photography, distortion is associated with focal distance and width camera's angle. Distortion correction is needed to be applied to photographs as it can cause image matching error. Equation (2) shows the distortion correction equation. The value x and y denote undistorted image point as projected by a camera. The value k_n is radial distortion coefficient and the value p_n is tangential distortion coefficient. Each photograph for SfM is translated to undistorted using this equation.

$$\begin{aligned} x_{corrected} &= x(1 + k_1r^2 + k_2r^4) + [2p_1xy + p_2(r^2 + 2x^2)] \\ y_{corrected} &= y(1 + k_1r^2 + k_2r^4) + [2p_2xy + p_1(r^2 + 2y^2)] \end{aligned} \tag{2}$$

2.4 Localization

For drawing virtual objects on a display, the system needs localization. Features of picture from the live video are extracted using SURF in real time, and are compared with the features of stored images used for SfM processing in the database. To achieve real time processing, some enhancement such as Locality Sensitive Hashing (LSH) is adopted. It is one of approximate nearest neighbor search techniques and is used for quick comparison of key features between live video image and all stored images in the database. Another enhancement technique is to limit the calculating area. Almost all pictures have the target structure in the center. Therefore, similarity matching between live video image and photographs for SfM is calculated using center of each image. Once the system calculates the scores and identifies the high score image, the position and posture of this image are registered as current position and posture. Then, using this position and posture, virtual objects are precisely rendered in the AR display.

2.5 Implementation

In this research, the main system was implemented using java. Additionally, open source software application and libraries were utilized. To reconstruct 3D virtual objects using SfM, Open MVG (Open Multiple View Geometry. Ver.0.7) was used. Trimble SketchUP (Ver.14.3.331), which is a 3D modeling software, was used for defining the coordinates of virtual objects. OpenCV (Open Source Computer Vision Library. Ver2.4.9), which is an image processing library, was used for extracting SURF features of captured real time camera stream and stored photographs, and for image matching. Finally, freeglut (Ver. 2.8.1), which is extended toolkit for OpenGL (Open Graphics Library), was used for drawing virtual objects.

3. VERIFICATION

To verify the applicability of the developed system for visualizing building information, a verification experiment was conducted outdoors. In this case study, the research building M3 of Osaka University Suita Campus was selected as the target building. The experiment was carried out on January 19, 2016 at 11:00 a.m. An optical see-through HMD, Moverio BT-200 with 1.2GHz of CPU (Texas Instruments OMAP 4460 Dual-core 1.2 GHz), 1GB of RAM, running android 4.0.4 was used in this experiment.

In this case study a 2D label was augmented to a part of M3 building as a prototype application. The virtual object is a CASBEE (Comprehensive Assessment System for Built Environment Efficiency) label, which is for assessing and rating the environmental performance of buildings and built environments in Japan. Using the prototype application, the user can check the condition of a building with the augmented CASBEE label. AR label includes information about overall performance evaluation, lifecycle CO_2, energy saving index, and heat island relaxation. The web chart on the label also show the primary assessment results.

Figure 3 shows the experimentation area. 3D point cloud of the building was created by SfM for the target building. In this case study, SfM used photographs with Exif (Exchangeable image file format) metadata captured by a digital camera. Figure 4 shows the route in which the photographs were taken with a red solid line. A hundred images (resolution 960 × 540) were taken and used for the SfM. 3D shape of M3 building reconstructed from SfM is presented in Figure 5. The drawing position of the label is then defined relative to the virtual model of M3 building. Figure 6 shows the label to be augmented on M3 building. Figure 7 shows the area that the augmentation results was designed to be displayed based on defined coordinates. In order to verify the system, a user initialized the system using an HMD. The position of the user is shown in figure 4. The system detected similar image between the live video stream and photographs for SfM. The system displays the most similar image to the user and receives confirmation input from the user's handheld in order to continue processing. The system then acquires camera parameters of the most similar image for registration and CASBEE label was augmented on the live camera feed of M3 building.

Figure 3. Experimentation area

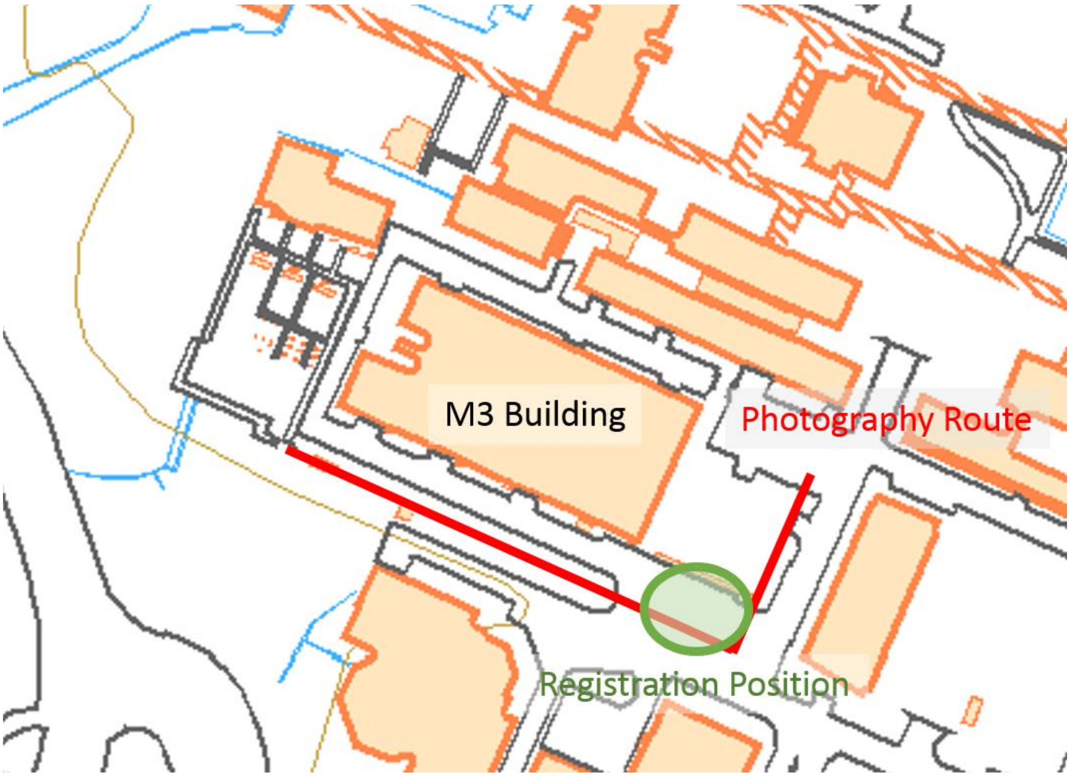

Figure 4. Photography route of SfM and registration point for experimentation

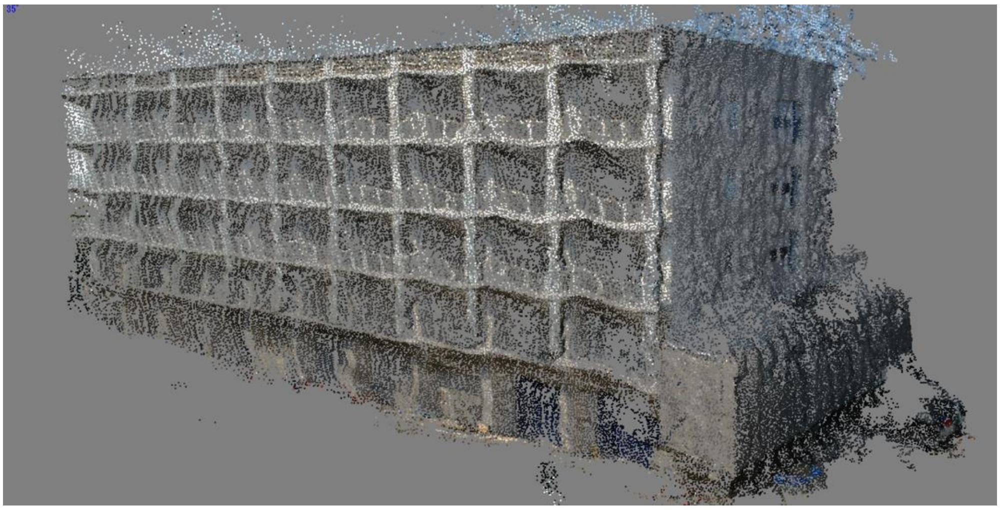

Figure 5. Point cloud reconstructed from SfM

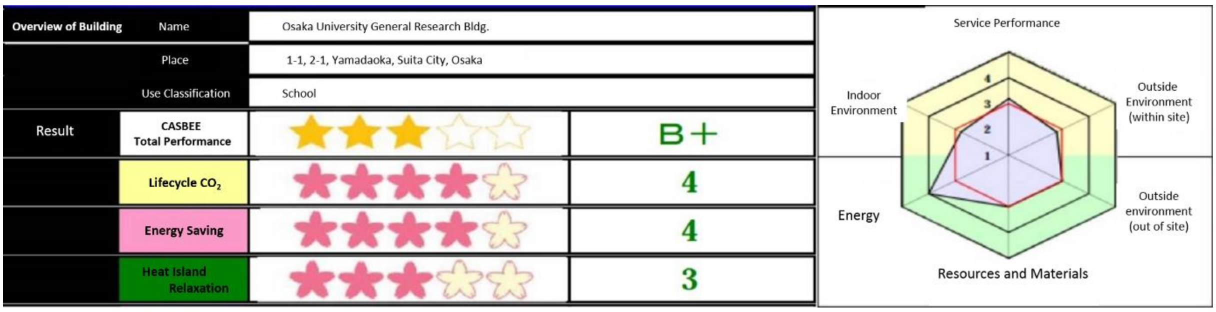

Figure 6. CASBEE label

Figure 7. Ideal result area of drawing a virtual object

4. RESULT AND DISCUSSION

This experiment demonstrated that the outdoor marker-less AR system can be used with the optical see-through HMD. Figure 8 shows a sample output of the prototype application. This figure contains the AR label augmented by the HMD's half mirror on the real environment scene that is seen by the user through HMD glasses.

Figure 8. Snapshot of the AR application

Although developed marker-less AR system for the optical see-through HMD was able to perform the registration process, visualization errors occurred. The average drawing position error for the AR label, with respect to the ideal position of the label, were 45.4 pixels in the horizontal direction and 25.2 pixels in the vertical direction. Additionally, the horizontal registration error ratio was 4.3% (horizontal registration error divided by the label's length) and the vertical registration error ratio was 10.8% (vertical registration error divided by the label's width). There are limitations and challenges in using the proposed system. As the developed system uses the same camera parameters as photographs used for SfM, it is only possible to use it at the route where the photograph for SfM were taken. Additionally, the registration can fail when there is no similar image in the database of the photographs even if the system is used in the same route (for example due to the difference between view angles of photographs). To solve these challenges, more route patterns can be used. Additionally, by increasing the number of photographs in each route the probability of finding similar photographs by the SfM will increase. Although taking a large number of photographs is costly, recent research shows that reconstruction by SfM can be achieved using video streams (Ventura & Hollerer, 2012).

Another challenge is the difficulty for the user to identify the AR object when there is low contrast between real environment and a virtual object which is drawn on a half mirror display with optical see-through HMD. A half mirror display is transparent and users can see real environment through this mirror. Hence, if the contrast is low, it is difficult for users to see the virtual object on the display or to grasp the relative position between real and virtual objects. Using a shade extension for HMD can improve the visibility when the contrast difference between a virtual object and real environment is low. However, this extension would not always improve the visibility. Therefore, new algorithms to automatically adjust the contrast between virtual objects and the real environment detected by the camera of HMD are required.

5. CONCLUSION

In this research, using optical see-thorough HMD, a marker-less AR system with local feature-based image registration technology and SfM which reconstructs 3D model by using photographs from several viewpoints was developed. The prototype of this proposed AR system has been verified in a case study. In the case study, registration and visualization of CASBEE label on M3 building located in Osaka University Suita Campus were tested and confirmed. The augmentation provided the user with the ability to see the information on the label.

ACKNOWLEDGEMENT

This research has been partly supported by JSPS KAKENHI Grant Number 25350010.

REFERENCES

Dellaert, F., Seitz, S., Thorpe, C., and Thrun, S. (2000). Structurefrom Motion without Correspondence, *IEEE Computer SocietyConference on Computer Vision and Pattern Recognition,* Vol.2, 557-564.

Karlekar, J., Zhou, S., Lu, W., Loh, Z,. Nakayama, Y. and Hii, D. (2010). Positioning, tracking and mapping for outdoor augmentation, *In mixed and Augmented Reality (ISMAR), 9th IEEE International Symposium on*, 175-184.

Reitmayr, G and Drummond, T. (2006). Going out: robust model-based tracking for outdoor augmented reality, *In Mixed and Augmented Reality, IEEE/ACM International Symposium on,* 109-118. IEEE, 2006.

Rolland, J. and Fuchs, Henly. (2000). Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization, *Presence: Teleoper. Virtual Environ.* Vol.9, 287-309.

Sato, Y., Fukuda, T., Yabuki, N., Michikawa, T. and Mmotamedi, A. (2016). A Marker-less Augmented Reality System using Feature-based Image Registration and Structure from Motion technologies for Building and Urban Environment, *Proceedings of the 21st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2016)*.(accepted)

Schubert, G., Schattel, D., Tonnis, M., Klinker, G. and Petzold, F. (2015). Tangible Mixed Reality On-Site: Interactive Augmented Visualisations from Architectural Working Models in Urban Design, *Computer-Aided Architectural Design, Communications in Computer and Information Science* 527, 55-74.

Ventura, J., Hollerer, T. (2012). Wide-area scene mapping for mobile visual tracking. *In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, 3-12

Wither, J., DiVerdi, S.Hollerer, T. (2009), Annotation in outdooraugmented reality, *Computers & Graphics 33*, 679-689.

Watanabe, S. (2011). Simulating 3D Architecture and Urban Landscapes in Real Space, *CAADRIA2011*, 261-270

Yabuki, N., Miyashita, K., Fukuda, T. (2011). An invisible height evaluation system for building height regulation to preserve good landscapes using augmented reality, *Automation in Construction, Volume* 20, Issue 3, 228-235.

Yabuki, N., Hamada, Y., Fukuda, T. (2012). Development of an accurate registration technique for outdoor augmented reality using point cloud data, *14th International Conference on Computing in Civil and Building Engineering.*

Zhang, Z. (2000). A flexible new technique for camera calibration, *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11), 1330-1334.