GAZE TRACKING CONTROL USING AN ACTIVE STEREO CAMERA

Masafumi NAKAGAWA*, Eisuke ADACHI, Ryuichi TAKASE, Yumi OKAMURA,
Yoshihiro KAWAI, Takashi YOSHIMI, Fumiaki TOMITA

National Institute of Advanced Industrial Science and Technology, 1-1-1, Umezono, Tukuba-city, Ibaraki, Japan -

(m.nakagawa, e-adachi, r-takase, y.okamura, y.kawai, tak-yoshimi, f.tomita)@aist.go.jp

Commission Ⅲ, WG Ⅲ/4

KEY WORDS: Active stereo camera, Object recognition, Segment based image matching, Gaze control, Real time processing, 3-D spatial data, Versatile Volumetric Vision

ABSTRACT:

The full automation of 3-D spatial data reference and revision requires spatial registration between existing spatial data and newly acquired data. In addition, it must be able to recognize an object’s shapes and behaviors. Therefore, the authors propose a real-time gaze tracking system capable of 3-D object recognition, in which an active stereo camera recognizes 3-D objects without markers. The real-time gaze tracking system was developed, and scenario-based experiments with the system were conducted. The results confirmed that our system could gaze and track moving objects successfully. Moreover, the proposed system achieves high-resolution 3-D spatial data acquisition and recognition, relative object behavior detection, and wide range covering.

1. INTRODUCTION

1.1 Background

Recently, semi automated procedures have been developed to achieve low-cost data handling in the field of 3-D Geographic Information Systems (GIS), such as 3-D urban data generation, 3-D urban data revision, and Intelligent Transport Systems. These procedures should be improved from semi automation to full automation for real-time data processing of real-time data.

The full automation of 3-D spatial data reference and revision requires spatial registration between existing spatial data and newly acquired data. In addition, it must be able to recognize an object’s shape and object’s behavior.

The optical flow algorithm is one of the traditional approaches to the detection of moving objects [1][2][3][4]. However, this approach has difficulty recognizing moving objects in images that contain occlusions, mainly because of the shortage of 3-D spatial information.

An image sensor has the advantage of high-speed data acquisition [5]. However, when a single camera makes an orbit around an object, the camera restricts available objects to simple shapes such as points and spheroids.

The Laser Identification Detection and Ranging (LIDAR) is also an effective sensor for detecting objects [6]. However, the low resolution of LIDAR requires manual registration for object recognition [7].

In addition, self-position estimation requires continuous 3-D information in a wide range environment. Usually, a fisheye camera has been used to acquire the wide range information [8]. However, the resolution of the camera is insufficient for generating precise 3-D spatial data.

1.2 Objective

The full automation of 3-D spatial data reference and revision requires the following capabilities to achieve spatial registration between existing spatial data and newly acquired data.

- high-resolution 3-D spatial data acquisition and recognition using image sensors without markers
- relative object behavior detection using temporal data, and
- wide range covering by a combination of camera translations and rotations

For local area surveys such as aerial photogrammetry from low-altitude flight, the authors believe that an active stereo camera is a suitable sensor for satisfying the above requirements. However, a gaze tracking procedure is necessary to realize the advantages of the active stereo camera. Therefore, we have developed a spatial registration system using an active stereo camera. In addition, a real-time gaze tracking system without markers is proposed in this research.

2. APPROACH

Here, we describe two cases of the gaze tracking procedures.

The first case is gaze tracking with known 3-D models such as existing 3-D urban data. When 3-D data have been prepared for an area, they can be used as reference data for the gaze tracking procedure. The known 3-D model could possibly have been prepared as a CAD model, generated via manual operations. Alternatively, the known 3-D model could be generated via a stereo matching procedure.

The second case is a gaze tracking without a known 3-D model. When no 3-D data has been prepared for an area, reference data must be prepared then and there, to be able to conduct the gaze tracking procedure.

This leads to three scenarios, described as follows.

- Scenario 1: Camera positioning via a known 3-D model.

The gaze tracking procedure is performed with a known 3-D model such as a CAD model (e.g. change detection by use of existing 3-D GIS data, such as camera positioning for autonomous robots).

- Scenario 2: Texture data acquisition via a known 3-D model.

The gaze tracking procedure is performed with a known 3-D model, generated through a stereo matching procedure (e.g. texture mapping for 3-D GIS geometrical data).

- Scenario 3: Gaze tracking of objects appearing in sequence data

The gaze tracking procedure is performed without a known 3-D model such as a CAD model or existing GIS data (e.g. pedestrian tracking or vehicle tracking).

A real-time gaze tracking system was developed using these scenarios. In addition, experiments were conducted for these scenarios to evaluate the performance of the system. Three approaches are described in this paper.

- Approach 1: The gaze tracking procedure with a known 3-D model (CAD model).

- Approach 2: The gaze tracking procedure with a known 3-D model (stereo matching procedure).

- Approach 3: The gaze tracking procedure without a known 3-D model.

3. THE REAL-TIME GAZE TRACKING SYSTEM

3.1 Concept of the real-time gaze tracking system

The system comprises an airborne stereo camera simulator system and a 3-D object recognition system. The airborne stereo camera simulator has five degrees of freedom, namely X-Y-Z translations and TILT-PAN axes of rotation. The 3-D object recognition system is applied to the real-time gaze tracking of objects. Continuous gaze tracking of objects is achieved by a combination of the simulator and the 3-D object recognition system.

A basic procedure in gaze tracking is to locate the object of interest in the center of a captured image, as follows. First, a model matching procedure uses the segment-based stereo spatial information of the object on the captured images. Then, the active stereo camera moves to catch the object in the centers of the captured images using the results of the previous model matching procedure.

3.2 Airborne stereo camera simulator

The hardware of the airborne stereo camera simulator system comprises a turntable and an active stereo camera hung on a crane (Figure 1).

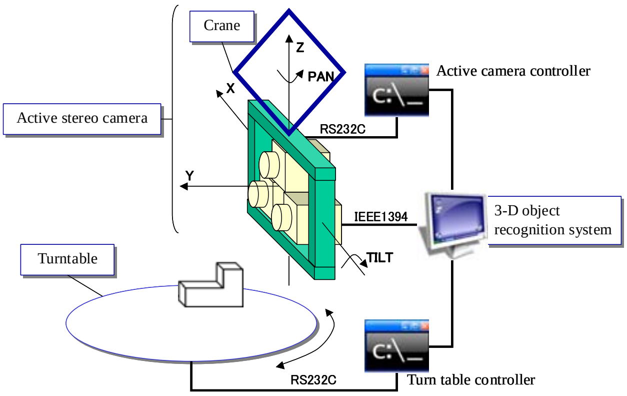

Figure 1. The airborne stereo camera simulator

Objects are placed on the turntable, which simulates horizontal rotation of the objects. The crane simulates vertical and horizontal translation of the objects. These rotation and translation data are transferred through controllers from the turntable and the crane to the 3-D object recognition system. Note that, these are simulated as relative rotation and translation parameters between the objects and the cameras.

The active stereo camera comprises three cameras mounted in one head. The head can rotate on TILT and PAN axes. The positions and orientations of the cameras are given from the angle of the turntable and the position of the crane.

3.3 3-D object recognition system

A block diagram of the gaze tracking procedure is shown in Figure 2.

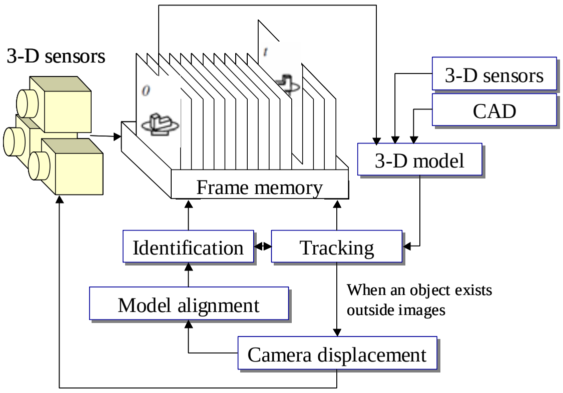

Figure 2. Block diagram of gaze tracking procedure

Traditional 3-D object recognition methodology cannot easily track moving 3-D objects in real time, because the processing time for 3-D object recognition is too great to track moving 3-D objects, even for a single stereo shot.

Here, we are developing Versatile Volumetric Vision (VVV) [9] technology for the 3-D vision system in the real-time gaze tracking system. The 3-D object recognition module of this system detects and localizes an object in the most recent frame of a 'frame memory,' which requires hundreds of megabytes of RAM. It is based on two continuous procedures, namely the rough matching and the precise matching of camera position estimation. These procedures perform the 'identification' in Figure 2. The '3-D model,' which is generated from CAD, or acquired via 3-D sensors, is used as the given data. 'Tracking' tracks object motion, frame-by-frame, within the stereo image sequence buffered in the frame memory.

Sequential stereo images must be captured and processed at high speed to track the objects in the gaze tracking procedure for camera position estimation. Here, Hyper Frame Vision technology [10] is used with a high-capacity frame memory. The computational time for the identification task generally exceeds one frame period. Therefore, the identification task requires substantial processing time. On the other hand, the tracking task takes less than one frame period. While waiting for the identification task to be completed, the stereo image sequence is buffered into the frame memory for the tracking task. Buffering the image sequence safely during the identification task can make use of the computational time difference between the identification task and the tracking task. Through memory sharing, independent tasks, such as recognition, tracking, and image viewing, are processed in parallel. By this means, the 3-D object recognition system provides real-time processing for the object tracking and gaze control.

In addition, when an object is outside the image field or beyond a threshold of camera view, a ‘camera displacement’ (see Figure 2) is executed to relocate the object of interest to the center of captured images by active stereo camera translation and rotation. The camera displacement outputs the repositioned object parameters for the images. Then, the object in the images is repositioned via model realignment. The repositioned object parameters are fed back from the camera displacement task to the identification task.

4. EXPERIMENTS

4.1 Overview

Three approaches were described in Section 2. Based on these approaches, three types of experiments were conducted to confirm the successful performance of gaze tracking using the airborne camera simulator.

Approaches 1 and 2 are the gaze-tracking procedures with a known 3-D model. The known 3-D models were generated via CAD in Approach 1. They were generated through stereo matching measurements in Approach 2.

Approach 3 is the gaze tracking procedure without a known 3-D model. Here, a 3-D model was generated via an image subtraction algorithm for reference data in the gaze tracking procedure.

4.1.1 Instrument

The airborne camera simulator developed is shown in Figure 3. Figure 4 shows details of the active camera. Figure 5 shows acquired images from the active camera. The simulator was designed along the lines of the discussion in Section 3.2.

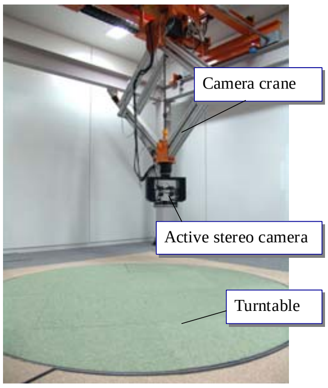

Figure 3. The airborne stereo camera simulator

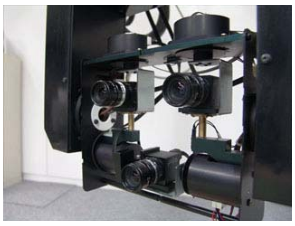

Figure 4. Active stereo camera

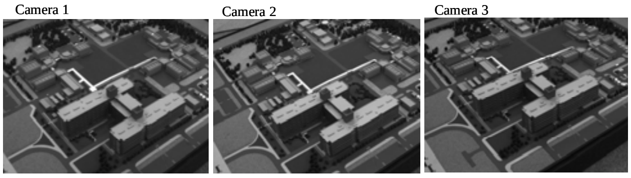

Figure 5. One acquired image shot

4.1.2 Material

Diorama building models were prepared for these experiments, as shown in Figures 6 and 7.

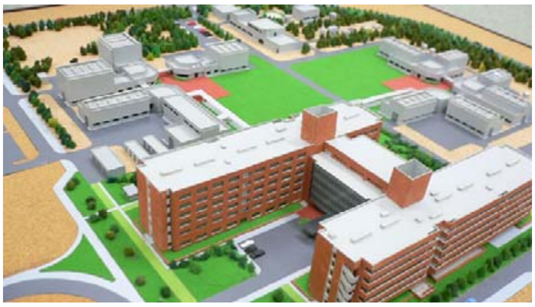

Figure 6. Building diorama model (1)

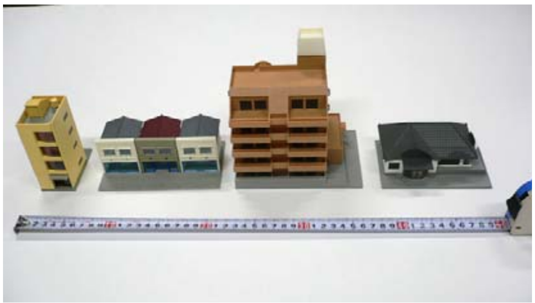

Figure 7. Building diorama models (2)

4.2 Prepared 3-D model

4.2.1 Approach 1: The gaze tracking procedure with a known 3-D model (CAD model)

A set of building diorama models and CAD model shown in Figure 8 are used as the material for this experiment.

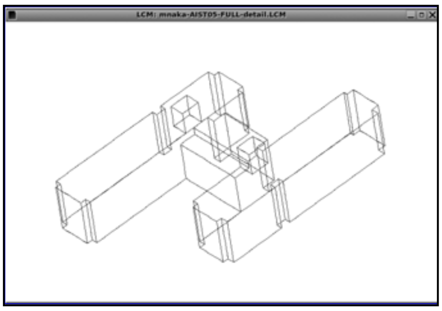

Figure 8. CAD model

4.2.2 Approach 2: The gaze tracking procedure with a known 3-D model (stereo matching procedure)

Four types of building diorama models are used as the material for this experiment.

3-D data for the building diorama models are generated through the stereo matching procedure as follows. First, stereo images of the building diorama models are captured from 18 directions, as shown in Figure 9. Next, 3-D segments of these shots are measured from each stereo image. Finally, each 3-D segment is merged to generate 3-D data without overlapping 3-D segments. The 3-D data generated from this procedure are shown in Figure 10.

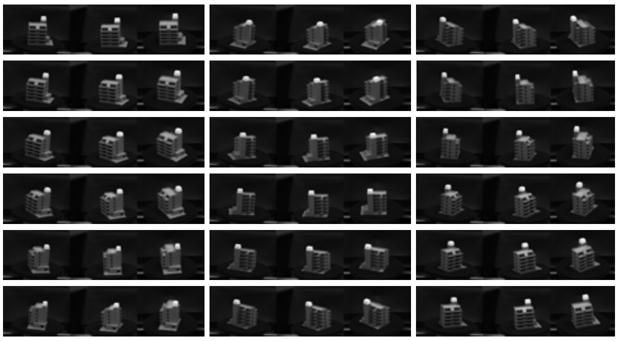

Figure 9. Captured stereo images from 18 directions

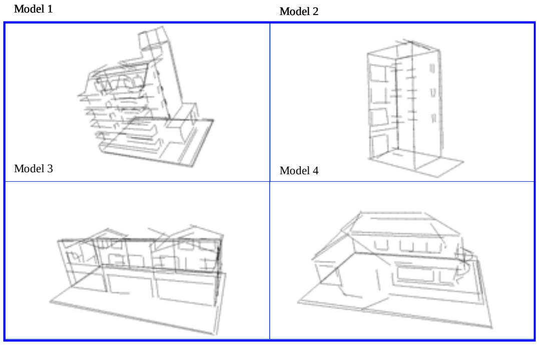

Figure 10. Prepared 3-D data for Approach 2

4.2.3 Approach 3: The gaze tracking procedure without a known 3-D model

Four types of building diorama models are used as the material for this experiment.

3-D data for the building diorama models are generated through the image subtraction algorithm, as shown in Figure 11.

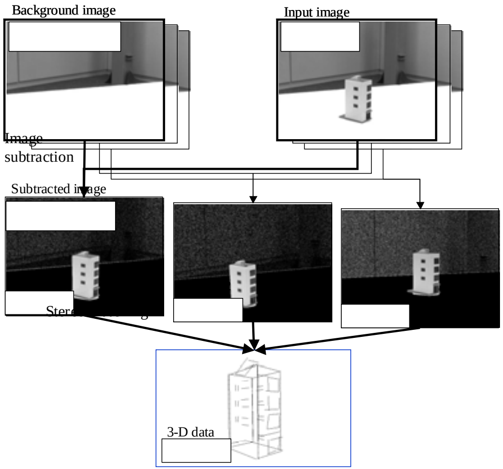

Figure 11. Image subtraction procedure

First, stereo images without the building diorama models are captured as background images. Next, stereo images including the building diorama models are captured as input images. Then, segments of the building diorama models are extracted through the image subtraction procedure, from the input images and background images. At this time, a median filter and a small segment filter remove minor noise, such as correspondence errors. Finally, the 3-D data for the extracted objects from one viewpoint are generated via stereo matching as shown in Figure 12.

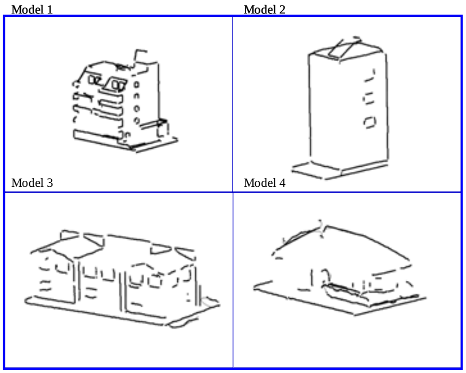

Figure 12. Prepared 3-D data for Approach 3

4.3 Gaze tracking

The experiment of gaze tracking is conducted as shown in Figure 13. An object rotates on the turntable, and the active stereo camera moves to track the object continuously. Sequence images are acquired in the gaze tracking procedure at 15 frames/sec. Sequence images at 180-frame intervals are shown in Figure 14.

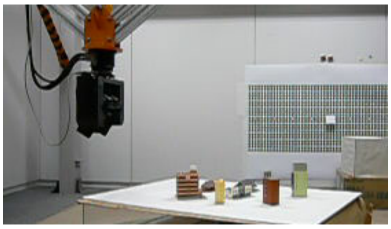

Figure 13. Gaze tracking experiment

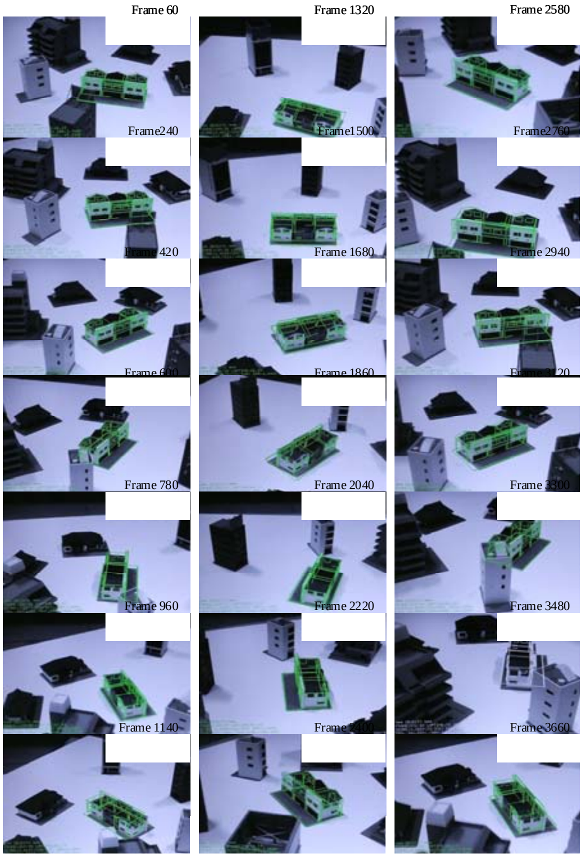

Figure 14. Sequence images in gaze tracking

5. DISCUSSION

Sequence images are acquired in the gaze tracking procedure at 15 frames/sec. Selected images from the sequence of images are shown in this section as input data and results. Geometrical accuracy is calculated using images captured from the various directions. Figures 15, 16, and 17 show the geometrical accuracy via 3-D models overlaid on the images captured from these directions. The variance of pixel error values in images is shown as the 2-D error2, and the variance of relative distance error values between the reference 3-D model and the measured results is shown as 3-D error2, in Table 1, 2, and 3.

5.1 Approach 1: The gaze tracking procedure with known 3-D model (CAD model)

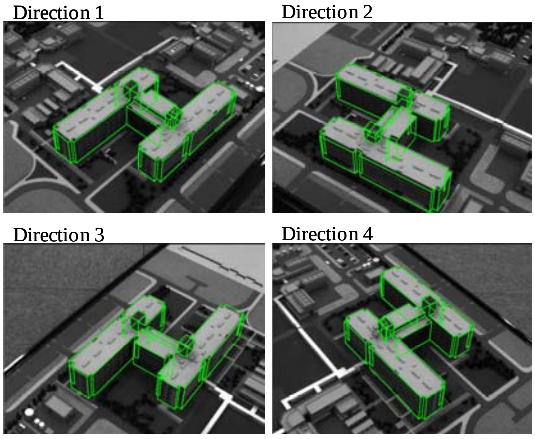

Figure 15. Results for Approach 1

Table 1. Results for Approach 1

Distance 50–60cm		
Direction	2D error2 [pix2]	3D error2 [mm2]
1	4.86467	4.83595
2	5.33922	5.07165
3	2.08338	5.18490
4	3.50135	4.46132
Average	3.94716	4.88846

5.2 Approach 2: The gaze tracking procedure with known 3-D model (stereo matching procedure)

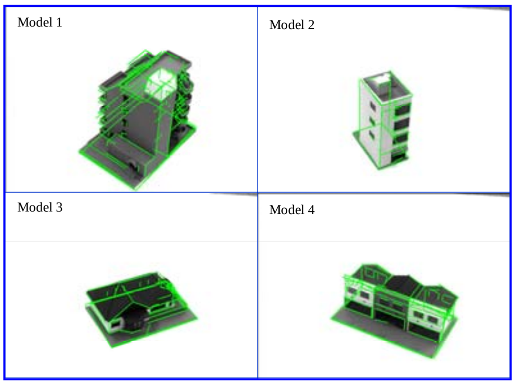

Figure 16. Results for Approach 2

Table 2. Results for Approach 2

Model1 Apartment

Distance 50–60cm		
Direction	2D error2 [pix2]	3D error2 [mm2]
1	3.57346	2.33406
2	4.20294	3.03302
3	5.75622	4.64325
4	4.02823	3.30154
Average	4.39021	3.32797

Model2 Simple building

Distance 50–60cm		
Direction	2D error2 [pix2]	3D error2 [mm2]
1	5.57506	4.39045
2	5.32057	4.10617
3	2.70112	3.11947
4	2.63687	2.16738
Average	4.05841	3.44587

Model3 Restaurant

Distance 50–60cm		
Direction	2D error2 [pix2]	3D error2 [mm2]
1	4.73232	3.50161
2	4.11710	2.29355
3	4.45813	2.73230
4	5.56429	3.46175
Average	4.71796	2.99730

Model4 Retail stores

Distance 50–60cm		
Direction	2D error2 [pix2]	3D error2 [mm2]
1	6.08988	2.59569
2	2.93348	3.31654
3	5.28145	3.66510
4	4.05963	2.75304
Average	4.59111	3.08259

5.3 Approach 3: The gaze tracking procedure without known 3-D model

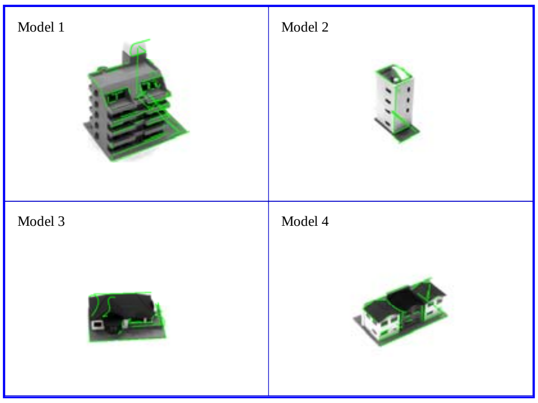

Figure 17. Results for Approach 3

Table 3. Results for Approach 3

Distance 50–60cm		
Model	2D error2 [pix2]	3D error2 [mm2]
1:Apartment	3.19023	3.38763
2:Simple building	1.49326	2.40126
3:Restaurant	1.39170	3.89351
4:Retail stores	0.82100	0.93149
Average	1.72405	2.65347

5.4 Overview

5.4.1 Geometrical accuracy

Tables 1, 2, and 3 show the variances of relative distance error values between the reference 3-D models and the measured results.

The average 3-D error2 in Table 1 is 4.88846. That is, the 3-D error in Approach 1 is 2.21[mm]. The 3-D error2 values in Table 2 are between 2.99730 and 3.44367. Therefore, 3-D errors in Approach 2 are between 1.73[mm] and 1.86[mm]. The average 3-D error2 in Table 3 is 2.65347. Therefore, the 3-D error in Approach 3 is 1.63[mm].

The accuracy of stereo matching measurement is approximately 1 mm in these experiments. Therefore, the geometrical accuracy of object recognition, from these results, is sufficient for the gaze tracking procedure in these experiments.

The 3-D errors in Approaches 1, 2, and 3 include not only object recognition errors but also 3-D model generation errors. The error in Approach 1 contains uncertainty about model geometry in CAD model generation. The errors in Approaches 2 and 3 depend on the accuracy of the stereo matching procedure.

5.4.2 Object recognition in gaze tracking

A 3-D model is used as reference data in object recognition. Therefore, an object can be recognized from its opposite side, as shown in Figures 18 and 19.

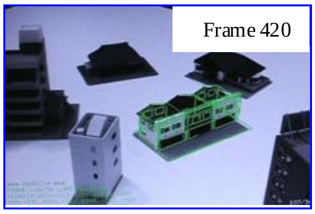

Figure 18. Acquired sheen

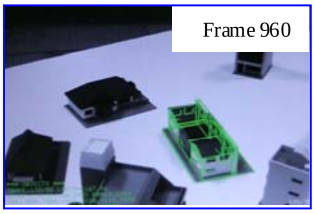

Figure 19. Opposite side

The object can be recognized even if occlusion exists, as shown in Figure 20. These additional results are shown in Figure 20.

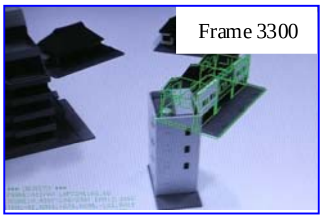

Figure 20. Occlusion

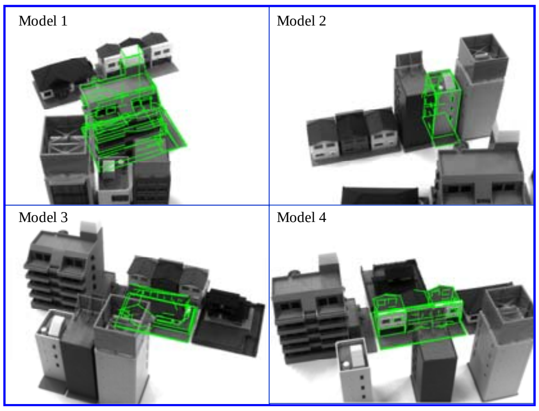

Figure 21. Object recognition with occlusion

When the area of occlusion is large, object recognition fails to track the object, as shown in Figure 22. However, the 3-D object recognition system recovers from object recognition error when the object appears again in the stereo images, as shown in Figure 23.

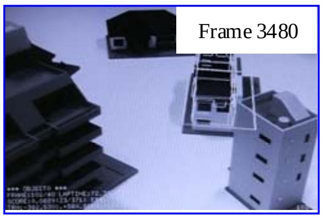

Figure 22. Tracking failure

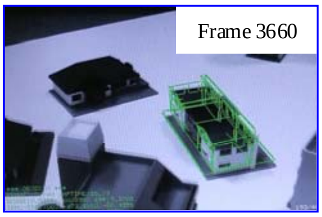

Figure 23. Tracking recovery

6. CONCLUSION

We have presented a real-time gaze tracking system with VVV, which enables its active stereo camera to recognize 3-D objects without markers. The combination of rough matching and precise matching in camera position estimation can gaze and track objects continuously. Moreover, Hyper Frame Vision can perform stable gaze tracking of moving objects in real time.

In this research, we have described three approaches, and conducted corresponding experiments. Two approaches used a gaze tracking procedure with a known 3-D model. The third approach used a gaze tracking procedure without a known 3-D model. From our results, we have confirmed that our methodology can gaze and track objects successfully. Moreover, the proposed system achieves high-resolution 3-D spatial data acquisition and recognition, relative object behavior detection, and wide range covering. We plan to improve the automation of actual GIS operations for 3-D map generation and 3-D map reference. In addition, we plan to mount an active stereo camera in autonomous navigation systems.

REFERNCES:

[1] B.K.P. Horn and B.G. Schunck. Determining optical flow. AI Memo 572. Massachusetts Institue of Technology, 1980.

[2] M. Proesmans, L. Van Gool, E. Pauwels and A. Oosterlinck. Determination of optical flow and its Discontinuities using non-linear diffusion. In 3rd Eurpoean Conference on Computer Vision, ECCV'94, Volume 2, pages 295-304, 1994.

[3] T. Camus. Real-Time Quantized Optical Flow. Journal of Real-Time Imaging, Volume 3, pages 71-86, 1997.

[4] McCane, B., Novins, K., Crannitch, D. and Galvin, B. (2001) On Benchmarking Optical Flow, Computer Vision and Image Understanding, 84(1), 126-143.

[5] Yoshihiro Nakabo, Idaku Ishii, and Masatoshi Ishikawa: Moment feature-based three-dimensional tracking using two high-speed vision systems, Advanced Robotics, Vol.17, No.10, pp.1041-1056, 2003

[6] Chris McGlone, with Edward Mikhail and James Bethel, Manual of Photogrammetry, 5th Edition, pp.629-636, 2004.

[7] Emmanuel P. Baltsavias, A comparison between photogrammetry and laser scanning, ISPRS Journal of Photogrammetry & Remote Sensing, 54(1):83-94, 1999.

[8] Nobuyuki Kita, Francois Berenger, Andrew Davison, Real-time Pose and Position Estimation of a Camera with Fish-eye Lens, Demonstration Session of International Conference on Computer Vision, 2003

[9] Tomita,F. Yoshimi,T. Ueshiba,T. Kawai,Y. Sumi,Y. Matsushita,T. Ichimura,N. Sugimoto,K. Ishiyama,Y., "R&D of versatile 3D vision system VVV", Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on, pp.4510-4516 vol.5, 11-14 Oct 1998

[10] Yasushi Sumi, Yutaka Ishiyama, Fumiaki Tomita, "Robot-vision architecture for real-time 6-DOF object localization", Computer Vision and Image Understanding Volume 105 - Issue 3, pp.218–230, 2007

[11] K.B. Atkinson, J.G. Fryer, Close Range Photogrammetry and Machine Vision, pp.78-104