Large-scale stereo for improvement of 3D measurement accuracy in gaze-observation

NAKAGAWA Masafumi, KAWAI Yoshihiro, TOMITA Fumiaki
National Institute of Advanced Industrial Science and Technology, Japan
m.nakagawa@aist.go.jp, y.kawai@aist.go.jp, f.tomita@aist.go.jp

Abstract

We propose a methodology that generates a large-scale stereo with a long theoretical baseline through a combination of small-scale stereos captured from various points. Moreover, we have confirmed that the stereo measurement accuracy is improved by improvement of the ratio between baseline and the distance between the camera and the object, even if the distance is large. We achieved this by camera pose and position estimation and 3D model-based tracking with 3D object recognition. We have also confirmed that our concept can improve the stereo measurement accuracy from a distant point by the verifying its accuracy, and in a verification experiment confirmed that our concept is valuable for actual stereo data such as aerial images. Thus, our approach allows a single stereo camera mounted on a moving object to achieve wide-range observation and high accuracy.

1 Introduction

Gaze observation is an approach to acquire data for all aspects of an object from various points using stereo cameras or laser sensors that are mounted on a moving object such as unmanned aerial vehicles or robots. Gaze observation is also an effective approach for 3D environmental data acquisition in a dense local area. Moreover, when we use a stereo camera mounted on a moving object, we require the stereo camera to cover a range as wide as possible in the depth direction with high accuracy. This requirement can be satisfied by the use of one stereo camera, which helps reduce platform weight.

Measurement accuracy with a stereo camera depends mainly on the baseline length, the distance from the camera to the objects, and the image resolution. When the image resolution is fixed, the measurement accuracy with a stereo camera depends on the ratio between the baseline length and the distance to the object, as shown in Figure 1.

Here, the baseline is defined as B, the distance from a camera to an object is defined as Z, and the side length of one pixel is defined as X. These parameters are described as $\frac{E}{Z} \cong \frac{2Z}{B}$. Therefore, the maximum error value in a stereo measurement is described as $E \cong \frac{2ZX}{B}$.

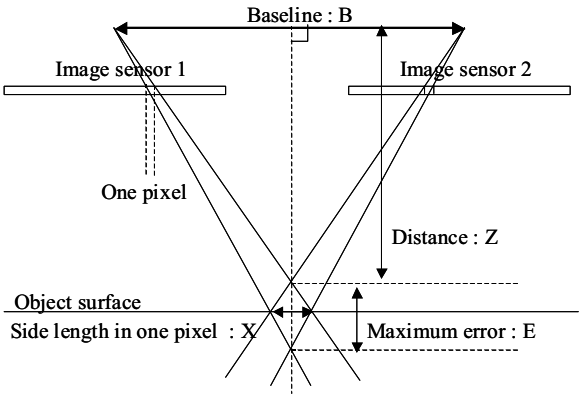

Figure 1: Stereo measurement accuracy

Although the baseline B should have a suitable value to conduct stereo matching procedures successfully, this equation shows that longer baselines give higher stereo measurement accuracy. This equation also shows that longer distances between the camera and the object give reduced stereo measurement accuracy.

2 Our concept and objective

In this research, we define a stereo image that is acquired with a short baseline as “small-scale stereo, ” and a stereo image that is acquired with a long baseline as a “large-scale stereo.”

Our concept and objective in this paper are shown in Figure 2. We propose a methodology that generates a large-scale stereo with a long theoretical baseline from a combination of small-scale stereos captured from various points. We also confirm that stereo measurement accuracy improves as the ratio between the baseline and the distance from the camera to the object improves, even if the distance to the object is large. We accomplish this by camera pose and position estimation and 3D model-based tracking with 3-D object recognition [1]. The advantages of this 3D model-based tracking compared with feature-point-based tracking for stability and continuity in gaze observation are described below.

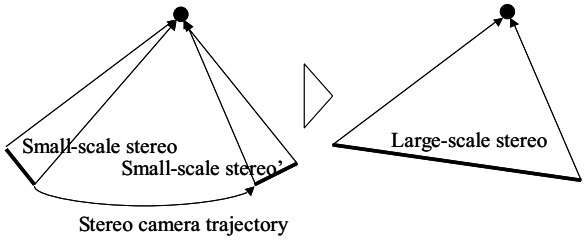

Figure 2: Small-scale stereo and large-scale stereo

2.1 Stability for tracking

The position and orientation of each camera are required to combine small-scale stereo images taken from various viewpoints. Two approaches for estimating camera position and orientation are feature-point tracking and 3D model tracking.

Generally, an optical flow algorithm is used for feature-point tracking [2], as shown in Figure 3. However, a rapid change in the lighting environment or a rapid change in camera motion often causes a failure in this tracking procedure, which results in a discontinuity in continuous images, therefore, tracking after such a failure is difficult.

On the other hand, the effects of rapid changes in lighting environment or camera motion are less severe for 3D model tracking than for feature-point tracking, because 3D model tracking observes the shape of an object, as shown in Figure 4. Therefore, the 3D model tracking approach can provide a stable tracking procedure.

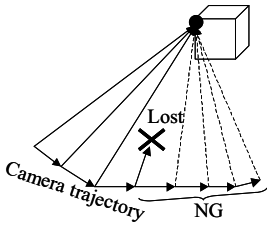

Figure 3: Feature-point tracking

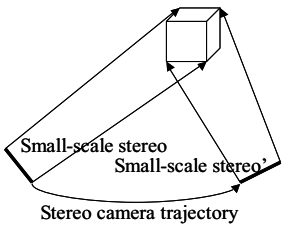

Figure 4: 3-D object tracking

2.2 Continuity for tracking

A feature-point tracking procedure can track a feature point whenever the procedure observes the feature point continuously, as shown in Figure 5. However, this procedure has the limitation that a feature point in a scene should exist near the position of the same feature point in the previous scene. Moreover, when an occlusion appears in continuous images, the tracking procedure becomes difficult. While we believe that there is a possibility of avoiding this problem with an algorithm such as intermittent feature-point update [3], its continuity for a tracking procedure is not adequate for gaze-observation.

On the other hand, a 3D model-tracking procedure can track an object whenever the procedure observes the shape of the object, as shown in Figure 6. Moreover, an object can be tracked even if there is a rapid camera motion change or the object is observed from the opposite side, because the object is tracked using object recognition. Therefore, 3D model tracking allows longer intervals in observation than feature-point tracking. Thus, strict continuity is not required in image capture. Moreover, even if a partial occlusion occurs in the acquired image, continuous tracking is still possible. Therefore, 3D model tracking has an advantage in continuity of tracking compared with feature-point tracking. As a result, we can conclude that 3D model tracking is suitable for gaze observation.

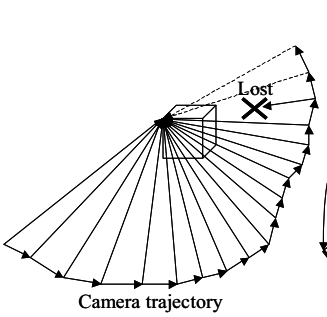

Figure 5: Feature-point tracking

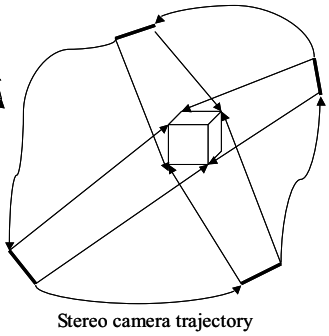

Figure 6: 3D object tracking

3 Methodology

The generation of a large-scale stereo taken from several small-scale stereo sets is shown in Figure 7. This figure illustrates the procedure that outputs the large-scale stereo set by combining the images and camera parameters of the small-scale stereo sets with different coordinate systems.

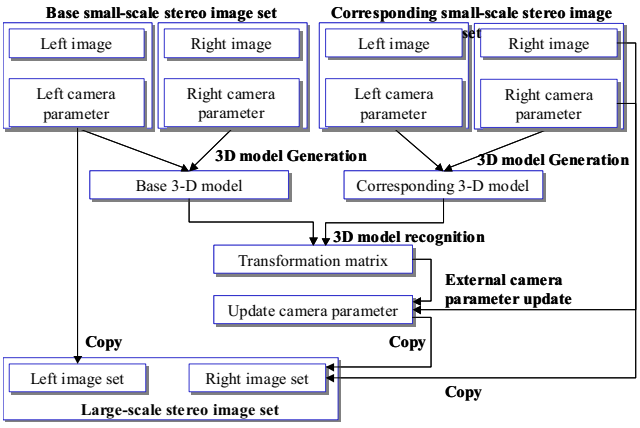

Figure 7: Methodology for large-scale stereo image generation

This methodology consists of the following three components:

- generation of 3D models using small-scale stereos;

- external parameter update;
- generation of a large-scale stereo.

3.1 Generation of 3D models using small-scale stereos

First, stereo images are captured to contain an object using calibrated small-scale stereos. Next, 3D measurements are performed on each small-scale stereo set. Then, 3D models are generated in each small-scale stereo coordinate system, and segment-based stereo is applied to the 3D measurements. However, measurement accuracy in each small-scale stereo may be low, because the baseline of the small-scale stereo is short compared with the distance from the camera to the object.

Here, we define the small-scale stereo that is the base data for the transformation as the "base small-scale stereo." We also define the small-scale stereo that is transformed data in the transformation described later, as the "corresponding small-scale stereo."

3.2 External parameter update

A transformation matrix to combine the above small-scale stereo sets is calculated with 3D object recognition.

The position and orientation of an object are expressed as a 4*4 transformation matrix as follows, where R is a 3*3 rotation matrix and t is a 3D translation vector that moves an object model.

$$T = \begin{pmatrix} R & t \\ 0\ 0\ 0 & 1 \end{pmatrix}$$

In other words, the recognition algorithm is a procedure for calculating T by comparing an object model and scene data, which is reconstructed using stereo vision. An object is recognized in two phases: initial matching and fine adjustment.

Then, the external camera parameters of the corresponding small-scale stereo are transformed from the corresponding small-scale stereo coordinates to the base small-scale stereo coordinates with this matrix T. When GPS data and IMU data are acquired, they may be used as initial values for the matrix calculation.

3.3 Generation of large-scale stereo

Using the above procedure, images in the same coordinate system are prepared. The large-scale stereo is generated using two images taken from these images. Values transformed in an external parameter-update procedure are used as external camera parameters. Even if the distance from the camera to an object is large, stereo measurement will be more accurate, because the ratio between the baseline and the distance from the camera to the object is improved.

In Figure 7, the two left images are taken from two small-scale stereos. However, our methodology allows a combination of left image and right image or a combination of both right images. Moreover, when more than three small-scale stereo sets are prepared, our methodology also allows large-scale stereo sets to contain more than three images.

4 Experiments

We conducted an accuracy verification that confirms that our transformation from small-scale stereo to large-scale stereo improves stereo measurement accuracy. We also conducted an operation verification experiment that confirms that our methodology can generate large-scale stereo images using small-scale stereo images without targets. Our excellent results are described below.

4.1 Accuracy verification

We have confirmed that the transformation from small-scale stereo to large-scale stereo improves stereo measurement accuracy, using small-scale stereo images containing nine targets.

The input small-scale stereo images containing the targets are shown in Figure 8. The baseline of this small-scale stereo is 20 cm, and the distance from the camera to the object is 300 cm. The stereo images were captured using the same stereo camera set at two points; the distance between them was 100 cm, measured manually. Therefore, the baseline of the large-scale stereo is 100 cm. The generated large-scale stereo images are shown in Figure 9.

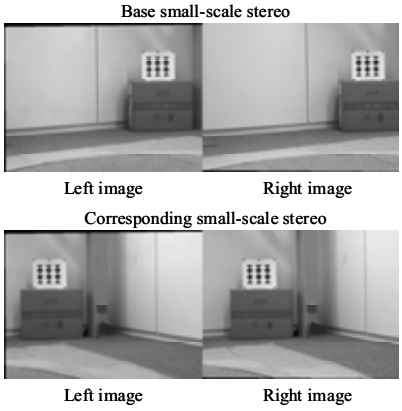

Figure 8: Input small-scale stereo

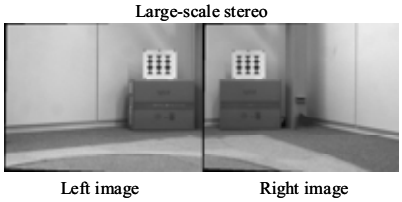

Figure 9: Output large-scale stereo

In this experiment, 12 vector points were measured between nine targets. A comparison of the stereo measurement values is shown in Table 1.

Table 1: Results of accuracy verification

Measured values by manual [mm]	Base small-scale stereo [mm]	Corresponding small-scale stereo [mm]	Large-scale stereo [mm]
79.30	77.99	78.13	79.31
79.30	76.19	76.23	79.58
79.30	76.25	76.31	79.39
79.30	78.51	76.35	79.61
79.30	76.28	76.63	79.68
79.30	77.79	77.09	79.48
59.50	57.97	57.93	59.64
59.50	58.18	57.23	59.66
59.50	57.95	57.21	59.60
59.50	60.06	59.32	59.70
59.50	58.02	57.09	59.56
59.50	58.01	58.38	59.60
Measurement accuracy [mm]	1.91	2.24	0.20

The measurement accuracy in small-scale stereo for one point was 1.91 [mm] (RMS), while the measurement accuracy in small-scale stereo for the other point was 2.24 [mm] (RMS). In contrast, the measurement accuracy in large-scale stereo was 0.20 [mm] (RMS). Therefore, we have confirmed that our transformation from small-scale stereo to large-scale stereo improves stereo measurement accuracy.

4.2 Operation verification

First, small-scale stereo images taken from an unmanned aerial vehicle were prepared for this experiment. These images contain a building without known targets. Next, 3D models were generated using several small-scale stereo images. Figure 10 shows these input images and 3D models.

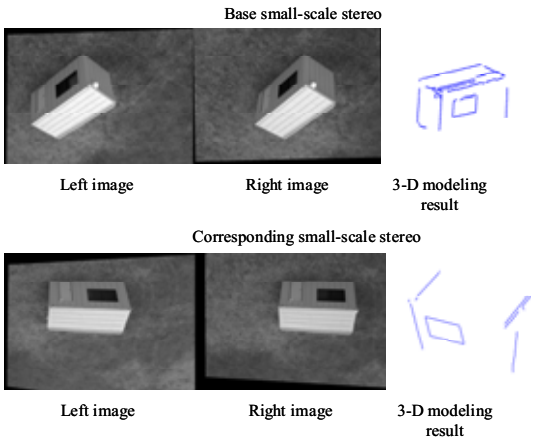

Figure 10: Input small-scale stereo

Next, the 3D object recognition procedure was conducted. Figure 11 contains the results in image space and 3D space.

Then, large-scale stereo images were generated from the small-scale stereo images. Figure 12 shows that an epipolar line exists in the correct state to confirm that the generated large-scale stereo can provide correct measurements.

These results confirm that large-scale stereo images can be generated from the combination of small-scale stereo images without targets.

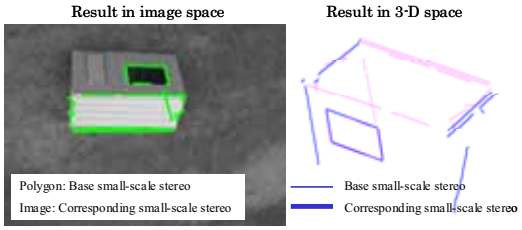

Figure 11: Result of 3D object recognition

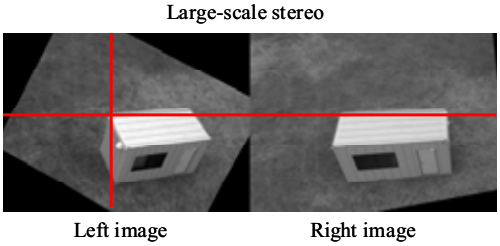

Figure 12: Output large-scale stereo

5 Conclusion

We have proposed a methodology that generates a large-scale stereo with a long theoretical baseline by combining small-scale stereos captured from several points. We have also confirmed that the stereo measurement accuracy is improved by increasing the ratio between the baseline and the distance from the camera to the object, even if the distance is large. In this paper, this concept was achieved by camera-pose position estimation and 3D model-based tracking with 3D object recognition.

We have confirmed that our concept can improve stereo measurement accuracy from a distant point, and our concept is applicable to actual stereo data such as aerial images. Therefore, our approach allows a single stereo camera mounted on a moving object such as an unmanned aerial vehicle or a robot to provide wide-range observation and high accuracy.

References

[1] Yasushi Sumi, Fumiaki Tomita.: "3D object recognition using segment-based stereo vision," Computer Vision — ACCV'98, Volume 1352/1997, 249-256, 1997.

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

[3] Buchanan, A, Fitzgibbon, A. : "Interactive Feature Tracking using K-D Trees and Dynamic Programming," Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1, 626-633, 2006.