On accuracy of 3D localization obtained by aligning 3D model with observed 2D occluding edges

Yasuyo KITA Nobuyuki KITA
Research Institute of Intelligent Systems
National Institute of Advanced Industrial Science and Technology (AIST)
y.kita{n.kita}@aist.go.jp

Abstract

In this paper, accuracy of the 3D localization obtained by aligning a 3D model with 2D observed occluding edges is discussed on. Aiming at accurate localization of a robot in a nuclear power plant, a method for aligning a 3D environmental model with an image observed by a camera mounted on the robot was proposed [1]. By effectively using the two-type predicted views which are calculated by a graphics system (eg. OpenGL etc), the method succeeded in robust alignment even though the scene consists of complicated occluding edges of pipes. However, accuracy of the 3D localization obtained by the alignment has not been yet enough examined. In this paper, some factors affecting the error in the 3D localization are investigated. The experimental results using both synthetic and actual data make it clear that the error in the focal length of the camera model causes relatively large translation error in the view direction. For the case the camera parameter is not known precisely, we propose to utilize two cameras to decrease the errors and show its effect.

1. Introduction

When the task of inspecting some environment is given to a robot, it is effective that the robot freely changes view points while freely moves around. Based on this philosophy, we have mounted a high-performance active camera head on a mobile robot aiming at autonomous inspection of nuclear power plants. Here, it is quite important to accurately know the position and pose of the camera head both to navigate the robot in a narrow space among the pipes and to carry out precise inspection. Therefore, we aim at visual feedback to correct inaccurate state of the camera head obtained from dead reckoning.

Since a 3D model of the environment surrounding a robot is given in our application, the 3D localization of the camera head can be done by aligning the 3D model with an observed image. Although some methods have been proposed for the 3D-2D alignment [2] [3], these are not directly applicable to our subject because of the complexity of observed occluding edges caused by many pipes in the plant. This difficulty was overcome by the method proposed in [1]. The method succeeded in robust 3D-2D alignment even in complicated scenes by effectively using the two-type predicted images which are calculated from the 3D environmental model by a graphics system (eg. OpenGL etc). However, accuracy of the 3D localization lead from the alignment result has not been yet enough examined.

In this paper, we investigate the accuracy of the 3D localization. By comparison of the results obtained by the 3D-2D alignment with manual measurements, we found relatively large translation errors in the view direction. The effect of some suspicious factors are examined through simulation using synthetic observed images. As a result, we show that inaccurate focal length of a camera model causes a big translation error in the view direction. Then, for the case that the focal length cannot be exactly known, we propose to utilize two cameras to decrease the errors. In Section 2, the basic scheme proposed in [1] is briefly explained. In Section 3, some factors affecting the localization error are analyzed. Then, the effect of the use of two cameras in decreasing the localization error is shown through the experiments using both synthetic and actual data in Section 4.

2 Basic Scheme of 3D-2D alignment[1]

Fig. 1 shows a scheme of our strategy for determining the position and pose of a camera by aligning the 3D model with occluding edges in an observed image [1]. Suppose that an image is observed by a camera whose initial position and pose are estimated (eg. data from dead reckoning). Because of the estimation error, the projection of the environmental model on the observed image is deviated as shown in Fig. 2a. The concrete procedures to correct the

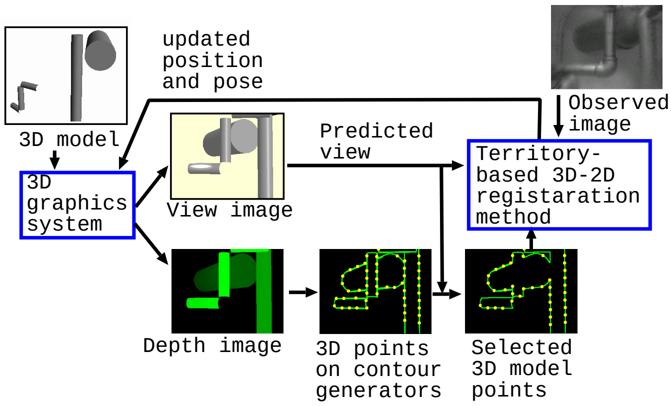


Figure 1. Scheme of detection of the position and pose of a camera using occluding edges

deviation are as follows:

1. The 3D model points corresponding to the observed edges are calculated from the 3D environmental model and the initial estimated state of the camera.

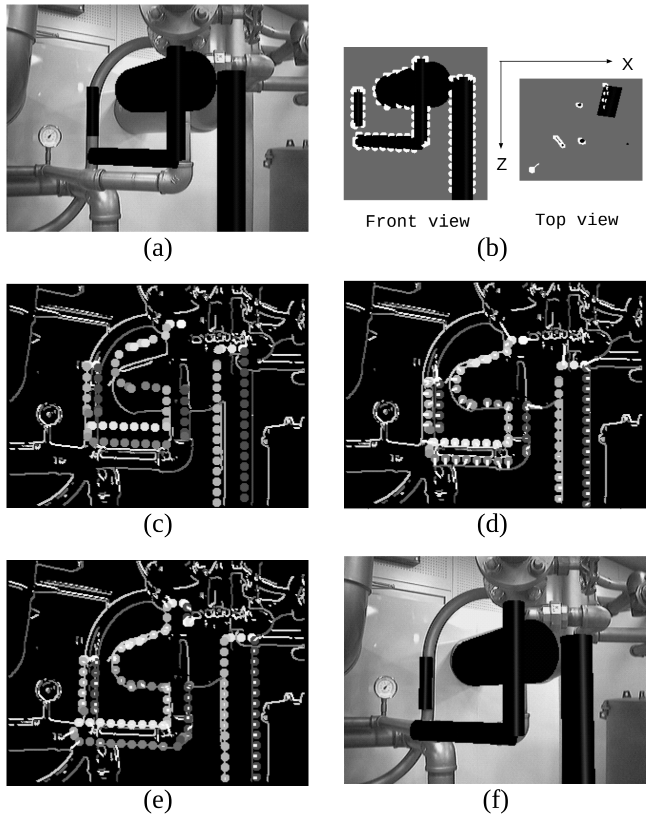
This process is done quickly by reading the 3D coordinates of the edge points of the depth image calculated by a graphics system (eg. OpenGL). White points in Fig. 2b show 3D model points calculated in such a way. In Fig. 2c, the model points are overlaid on the observed edge image calculated with Canny operator[4]. The grey levels of the model points and the edges illustrate their directional attributes which are classified into eight directions.

2. Observed edge points corresponding to the 3D model points are determined based on the closeness on the observed image.

Since a little change in camera angle causes a big translation in the image, only at the first time, before the 3D-2D matching, the projected 3D model points are two dimensionally translated on the image to the position where the model points overlap best on the observed edges with the same direction attribute. Fig. 2d shows the position after this initial translation. Territory-based 3D-2D matching[5] which uses anisotropic search restricted regions obtained from the projected shape of the model enables a high ratio of correct pairs. White lines connecting the model points and edges in Fig. 2d show the 3D-2D point correspondences obtained by the territory-based matching.

3. The current position and pose of the camera is renewed to satisfy the 3D-2D point correspondences.

The 3D state of the camera is calculated based on least-squares estimation at the present. Because of the usage of lots of pairs and a high ratio of correct pairs, the deterioration by wrong pairs looks not so big. When we aim at more accuracy, it is a way to adopt some robust estimation method[6].



(a) (b) (c) (d) (e) (f)

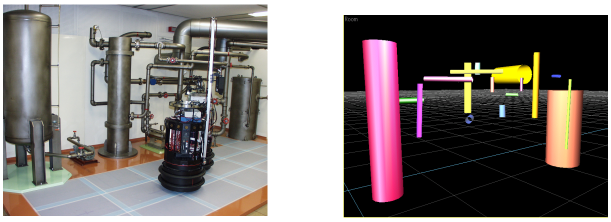
Figure 2. Example of processes of 3D-2D alignment: (a) projection of the model at the initial state; (b) front and top views of the 3D model points (white points); (c) projection of the 3D model points on the observed edge image; (d) projection of the 3D model points after initial 2D translation; (e) projection of the 3D model points after convergence; (f) projection of the 3D model after convergence.

At the new state, the same processes except the 2D translation on the image are iterated until the camera state converges. Fig. 2e, f shows the model projection after convergence. The computational time is less than a few sec (Pentium II(333MHz)). We plan to use this visual feedback not all the time but when the accurate localization is required for some specific tasks. The process is enough quick for the purpose.

3 Investigation on accuracy in 3D localization

3.1 Accuracy in 3D localization

We examined accuracy of the 3D localization calculated from the 3D-2D alignment method by comparing with manual measurements. Fig.3a shows our experimental environ-



(a) (b)

Figure 3. Experimental environment: (a) plant-mockup; (b) its partial models consisting of 17 cylinders.

ment, a plant-mockup. A robot with an active camera head moves around in the environment. 17 pipes in this environment are selected and modeled with cylinders in OpenGL as shown in Fig.3b. The 3D world coordinate system is defined as shown in Fig. 2b so that the x and z axes lie in the horizontal floor face; the y axis completes the left-handed coordinate system, and is in the vertical direction.

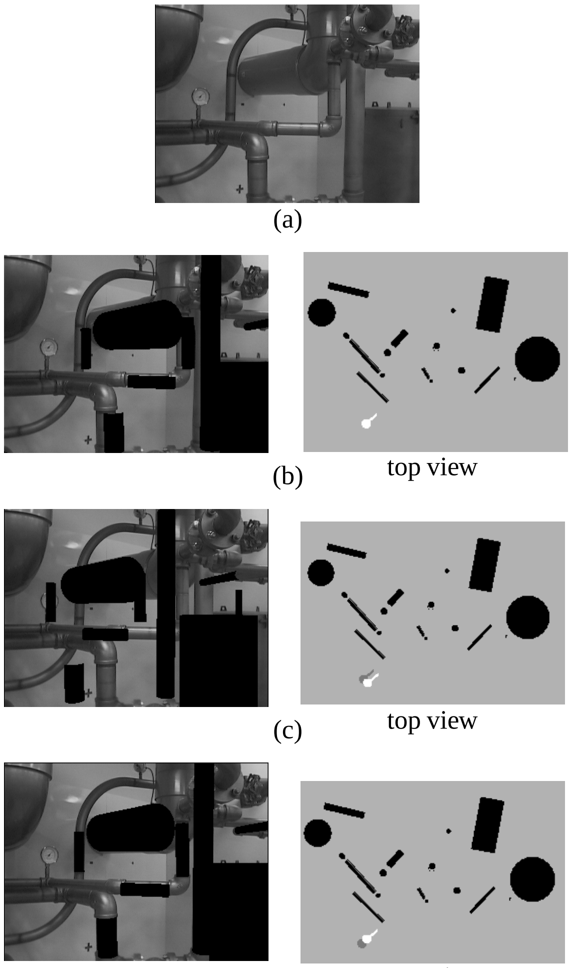
Fig. 4a shows an example of the images observed by a camera mounted on the robot. The position and pose of the camera head was manually measured with great care and illustrated in the top view of Fig. 4b: the white circle and the white line sticking out from the circle illustrate the position and the view direction respectively. The accuracy of the manual measurement is about $\pm$5mm in translation and $\pm$3 degrees in rotation. Because of this slight error, the projection of the model at the state shows a little deviation from the observed image as shown in Fig. 4b.

We intentionally add some errors to the camera state and use it as the initial estimate. Fig. 4c shows the projection of the 3D model when giving the camera state after adding (50,0,50) mm translation and 5 degree rotation around the y axis to the measured state. In the top view, the white circle shows the current camera position, while the gray circle overlapped by the white circle shows the measured state. Fig. 4d shows the result after correcting the camera state by the method described in Section 2. The model is well aligned with the image. Nevertheless, as shown in the top view, the translation error occurred mainly in the view direction, which is about 90 mm.

We have done similar experiments using more than 10 images observed at various locations. In all the experiments, 3D models are well aligned with observed images and 3D localization is converged. This showed the robustness of the method in such a complex scene. However, the translation error in the view direction appears in all cases.

3.2 Analysis of error factors

The factors causing the 3D localization errors can be counted up as follows:

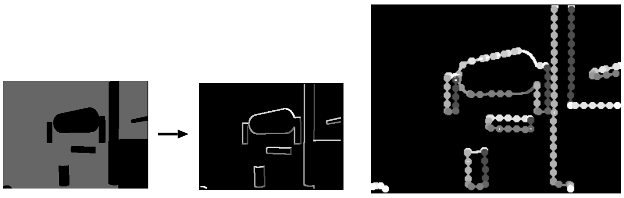


(a)
(b) top view
(c) top view
(d) top view

Figure 4. Example of localization result: (a) observed image; (b) measured state; (c) initial state; (d) result

1. Pixel quantization
2. Camera internal parameters
3. Inaccuracy of the 3D models
4. Wrong 3D-2D correspondences

The accuracy of our 3D model is based on the manual measurement and about $\pm$5mm in translation and $\pm$3 degrees in rotation. Since we calculate the camera state from the 3D-2D corresponding pairs based on least-squares estimation at the present, the wrong correspondences left at the final state deteriorate the accuracy. However, these errors should produce random errors in the 3D localization. From the observation of the clear tendency for the error to be translation in the view direction, we focus on the two suspicious factors, pixel quantization and the focal length of the camera internal parameters. By treating the focal length



(a) (b)

Figure 5. Experiment for investigating the effect of pixel quantized error: (a) calculation of synthetic observed edges; (b) projection of model points after convergence

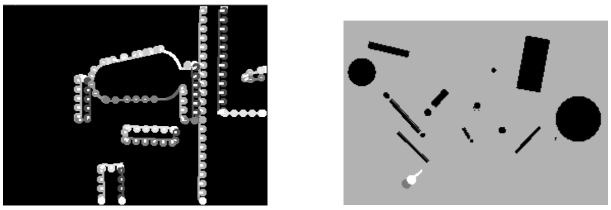


Figure 6. Effect of error in the focal length

in pixels, we include the effect of the error of the pixel aspect ratio in the factor. Here, we assume the horizontal and vertical pixel aspect ratios are the same, that is, the CCD's one pixel element is a regular square.

3.3 Effect of pixel quantization

To analyze the effect of the pixel quantization, we conducted the following synthetic experiments. First, the observed edges are synthetically calculated by projecting model data as shown in Fig. 5a. The 3D-2D alignment method is applied to this synthetic edge image. The correct state is given as the initial estimate so that the transformation from the initial state directly shows the effect of pixel quantized error. Fig. 5b shows projected model points converging on the synthetic observed edges.

From the results of this experiment, it was found that the error in the 3D location caused by the quantized error is small: the translation and rotation errors are about 0.6 mm and 0.03 degrees. This accuracy is supported by the fact that the method uses lots of 3D-2D corresponding pairs which distributed in a whole image (in this case, 122 pairs). Actually, if only nine pairs of the 3D-2D correspondences are used, the translation and rotation errors become about 11 mm and 0.4 degrees. Additionally, the effect of using distributed pairs was assured by the observation that the translation and rotation errors become about 4 mm and 0.08 degrees if we use the 3D-2D pairs only from two pipes, one vertical pipe and one horizontal pipe in the center of the image.

3.4 Effect of inaccurate focal length

Through similar synthetic experiments, the effect of inaccurate focal length of a camera model was examined. This time, synthetic observed edge images are calculated from the model projection obtained with the focal length which is slightly different from the camera model of the 3D-2D alignment method.

Fig. 6 shows a result when giving a longer focal length for producing synthetic views. The localization of the camera becomes closer in the view direction just as we experienced in the actual experiments. When giving a shorter focal length for synthetic views, the location deviated further in the view direction. In the situation in Fig. 6, the magnitude of the translation error in the view direction is about -36mm per 6.5 pixel (about 0.1 mm) error in the focal length.

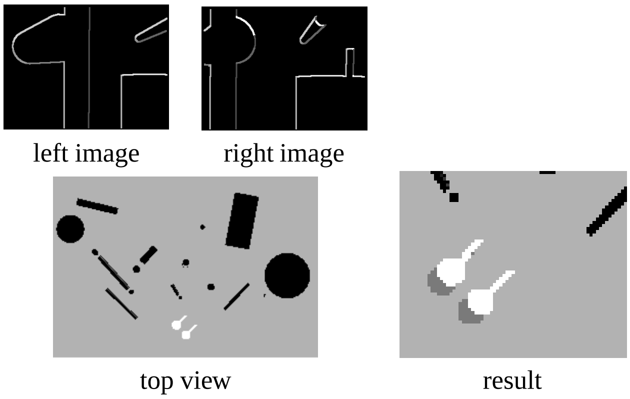
After this observation, we carefully measured the angle of the field of view of the actual camera to calculate the focal length. We found it is actually 48 degrees, although we had used the focal length corresponding to 50 degree angle of the field of view. In the case of data in Fig. 4, translation error in the view direction is decreased from 90 mm to 4mm by correcting this camera parameter.

4 Usage of two cameras for compensating the error

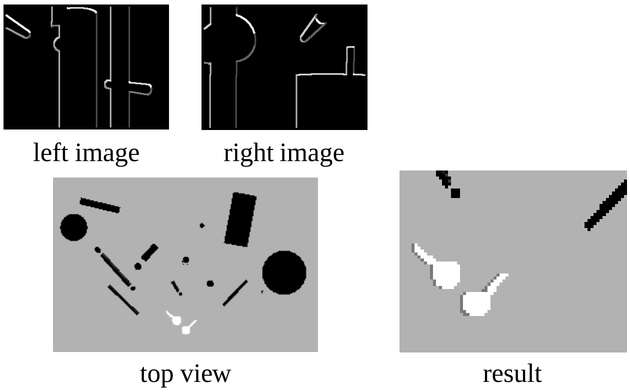
Although the translation error in the view direction can be decreased by using accurate focal length, it is sometimes difficult to know the accurate values especially when a robot need to change the camera focus and/or zoom during a sequential task. Therefore, in this section, we think about a way to compensate the error. From the point that the translation error is in the view direction, additional observation by another camera which has the view direction perpendicular to that of the first camera is thought to be effective to decrease the error. Actually, this addition is easy in our application since an active stereo camera head having controllable vergence is mounted on the robot for carrying out various tasks.

First, we synthetically simulated the effect of using two cameras. Fig.7a shows the left and right synthetic edge images of the parallel stereo cameras which are set as shown in the top view. Here, the synthetic edges are calculated using the focal length corresponding to 48 degree angle of the field of view.

The method is applied to the synthetic data using an inaccurate focal length, the focal length corresponding to 50 degree angle of the filed of view. The resultant translation and rotation errors were examined while panning the left camera outward at the same position. Fig.7b shows the left



(a) pan angle of left image = $0°$



(b) pan angle of left image = $90°$

Figure 7. Experiments on effect of using two cameras (synthetic data)

and right image and the camera state after 90 degree panning. The results are summarized in Table 1. The second line of Table 1 shows the result when using the right image only. From the third line of Table 1, the result when using the left image only and that when using the two images simultaneously are shown alternatively. Combining the 3D-2D correspondences obtained from two (or more) observed images can be done as shown in [5]. The cause of the dispersion of the translation error magnitude during panning is that the observed objects are changed by the panning. As shown in Table 1, the translation errors in the view direction are always improved by using two images except parallel stereo camera setting (pan angle = $0°$). The translation errors tend to become smaller when the angle between the view lines of the two cameras gets close to the right angle. This effect clearly appears in the difference between the the results of Fig.7a and b.

Next we examined the effect of two cameras by using actual images in Fig.8. Fig.8a shows the results using the two images of parallel stereo cameras. The translation error in the view direction was 45.4 mm. Fig.8b,c show the results using the images taken by the two cameras set so that the

Table 1 Localization errors caused by inaccurate focal length.

angle of the view directions	used image for localization	magnitude of total trans. error (mm)	trans. error in the view direction (mm)	rot. error (deg.)
-	R	58.2	-55.7	0.7
0	L	76.1	-74.6	0.6
0	L,R	59.7	-58.8	0.4
20	L	33.7	-32.6	0.5
20	L,R	29.1	-29.1	0.2
40	L	30.1	-29.6	0.7
40	L,R	15.8	8.7	1.1
60	L	19.7	-16.0	1.3
60	L,R	11.5	-10.2	0.7
80	L	59.0	-57.7	0.6
80	L,R	20.2	4.4	0.6
90	L	55.5	-55.0	0.2
90	L,R	16.5	11.7	0.7

angle of their view directions becomes 45 degrees. If we use the right image only to calculate the 3D location, the resultant position was deviated in the view direction of the right camera as shown in the top view of Fig.8b. The translation error in the view direction is 42.0 mm. As a result, the model projection on the left image at the resultant state deviated from the observed image. On the other hand, in the case we use the two images simultaneously, the translation error in the view direction are decreased to 16.0 mm as shown in Fig.8c.

5 Conclusion

In this paper, we investigated accuracy of the 3D localization obtained by aligning a 3D model with observed 2D occluding edges. Simulation using inaccurate focal length of the camera model clarified that its error brings relatively large translation error in the 3D localization. For the case the camera parameter is not known precisely, we proposed to utilize two cameras to decrease the errors. The experiments using both synthetic and actual data showed that the use of two cameras improves the localization accuracy when the angle of their view directions becomes closer to the right angle.

From the experimental results, the method seems to offer 3D localization accuracy similar to careful manual measurements, that is about $\pm$5mm in translation and $\pm$3 degrees in rotation. These values are enough for the purpose of the robot navigation in narrow spaces of the plant. Aiming at

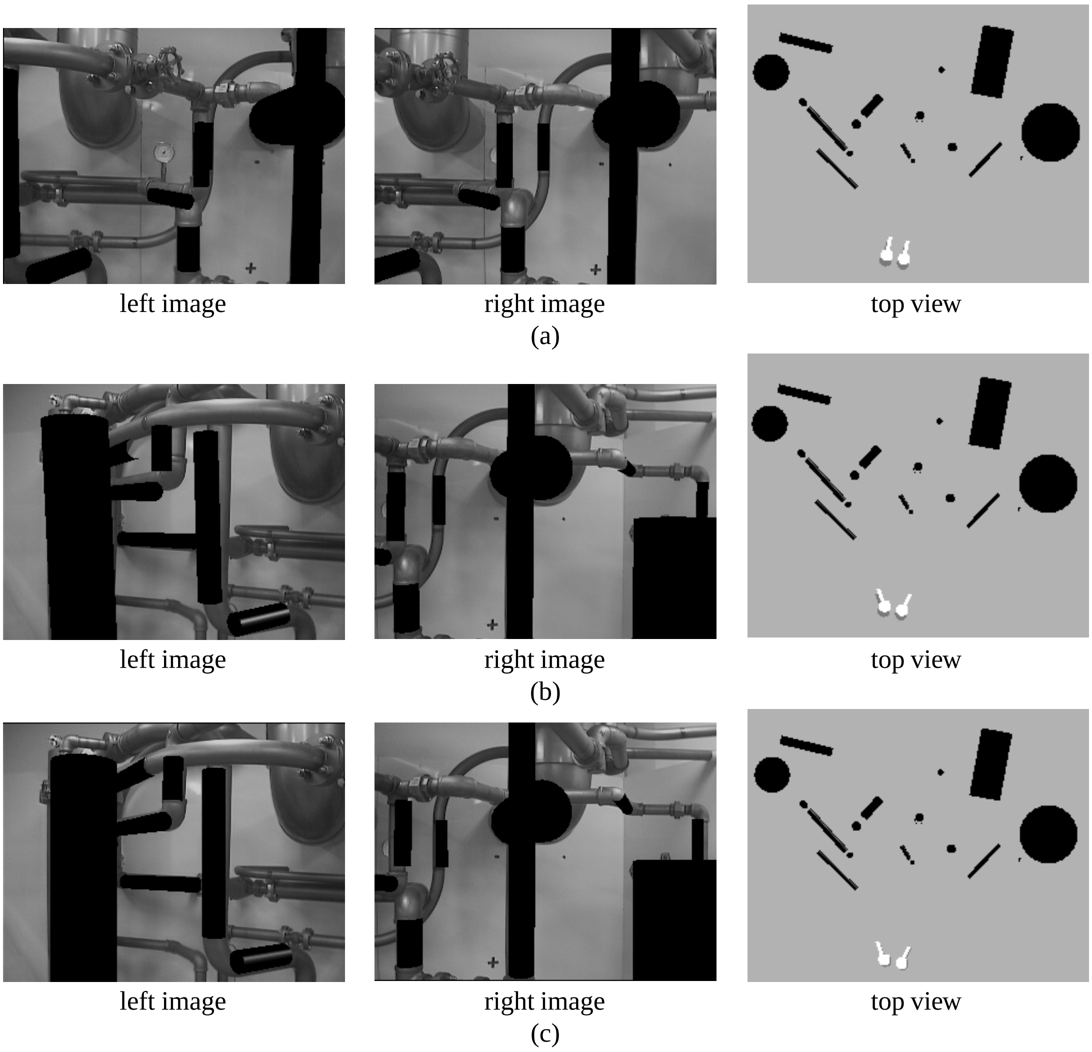


Figure 8. Experiments on effect of using two cameras (actual data)

applying to the tasks which requires more accuracy[7], we will investigate the effects of more various factors noted in Section 3.1.

Acknowledgments

We are thankful to Dr. Kazuo Tanie, Dr. Shigeoki Hirai, Dr. Takashi Suehiro and the members of Research Institute of Intelligent Systems.

References

[1] Y. Kita and N. Kita: “Position and pose detection of active camera-head in a nuclear power plant”, In *Proc. of Int. Conf. on Intelligent Robots and Systems (IROS 2000)*, pp. 1872–1879, 2000.

[2] S. Lavallée, and R. Szeliski: “Recovering the position and orientaion of free-form objects from image contours using 3D distance maps”, *IEEE trans. Pattern Anal. & Mach. Intell.*, **17**, 4, pp. 378–390, 1995.

[3] J. Feldmar, G. Malandain, N. Ayache, S. Fernandez-Vidal, E. Maurincomme and Y. Trousset:“Matching 3D MR Angiography Data and 2D X-ray Angiograms”, In *Proc. of CVRMed-MRCAS’97*, pp. 129–138, 1997.

[4] J. Canny: “A Computational Approach to Edge Detection”, *IEEE Trans.Pattern Anal. & Mach. Intell.*, Vol. 8, No. 6, pp. 679–698, 1986.

[5] Y. Kita, D. L. Wilson, J. A. Noble and N. Kita:“A quick 3D-2D registration method for a wide-range of applications”, In *Proc. of 15th International Conference on Pattern Recognition*, pp.981–986, 2000.

[6] M. D. Wheeler: “*Automaic modeling and Localization for object recognition*”, PhD thesis, Carnergie Mellon University, 1996.

[7] N. Kita, Y.Kita and Y. Hai-Quan: “Archiving technology of the plant inspection images captured by mobile active cameras”, In *Technical Report 2001-CVIM-129-5, Information Processing Society of Japan*, pp. 33–38, 2001 [In Japanese].