3-D Position Detection of Partner Robot using SURF Descriptor and Voting Method for Indirect Cooperation Between Multiple Robots

Toyomi Fujita and Kento Yamada
Department of Electronics and Intelligent Systems, Tohoku Institute of Technology, Sendai 982-8577, Japan

Keywords: Robot Vision, Cooperation by Observation, SIFT (Scale-Invariant Feature Transformation), SURF (Speeded Up Robust Features), Stereo Vision.

Abstract: In some practical work by robots, it may happen that a working robot can not detect a target object for handling due to a sensor occlusion. In this situation, if another cooperative robot observes the working robot with the target object and detects their positions and orientations, it will be possible for the working robot to complete the handling task. Such behavior is a kind of indirect cooperation. This study considers a method for such an indirect cooperation based on an observation by the partner robot. The observing robot will be able to perform such a cooperation by obtaining feature points and corresponding points on the working robot with hand and the target object from multiple captured images, then computing 3-D positions of the targets and motion of the hand. In this study, we mainly focus on 3-D position detection of the working robot and try applying SURF (Speeded Up Robust Features) descriptor and a voting method for detecting the feature points and corresponding points. The 3-D position of the working robot is then computed from these corresponding points based on stereo vision theory. Fundamental experiments confirmed the validity of presented method.

1 INTRODUCTION

Cooperation by multiple robots is effective to accomplish tasks. Multiple robots are able to complete a task even when one robot is not able to perform it by itself in some complicated environment. For example, let us consider a situation in which a mobile working robot that has a camera and a manipulator can not detect a target object to handle due to an occlusion by its arm for manipulation. In such a situation, if another mobile robot that has a camera observes the working robot, which is its partner, with the target object and detects their positions and the hand motion for manipulation, it can assist the handling of the working robot indirectly by sending the information to the working robot.

Such behavior is a kind of *indirect cooperation*. This study considers a method for such an indirect cooperation by the observing robot. The observing robot will be able to perform such a cooperation by obtaining feature points and their corresponding points on the working robot with hand and the target object from multiple images captured at different positions, then computing their 3-D positions and motion of the hand. In this study, we mainly focus on 3-D position detection of the working robot. We try applying SURF (Speeded Up Robust Features) descriptor (Bay et al., 2008) and a voting method for detecting the feature points and corresponding points. 3-D position of the working robot is then computed based on stereo vision theory. Fundamental experiments are conducted to confirm the validity of presented method.

2 POSITION DETECTION OF WORKING ROBOT

2.1 Robot Detection

Before detecting the 3-D position of the working robot, the observing robot needs to detect it from captured images by a mounted camera. In order for the detection, the observing robot compares two sets of feature points on the working robot: one set is obtained from an image captured in advance and another set is obtained from an image currently observed. Two points are picked from each set and they are corresponded if they have close feature values each other. If the number of all possible corresponding points is over than a threshold, we can consider that there is

Fujita, T. and Yamada, K.
3-D Position Detection of Partner Robot using SURF Descriptor and Voting Method for Indirect Cooperation Between Multiple Robots.
DOI: 10.5220/0006008905220525
In *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2016) - Volume 2*, pages 522-525
ISBN: 978-989-758-198-4

the working robot in the view of the image currently observed. Then the area with which corresponding points cover can be extracted as that of the working robot in the image.

In this study, we propose a voting method for efficient detection. The captured image is divided by several square regions. The working robot is recognized if there are matched regions in which the number of corresponding points is more than a threshold given in advance. The matched regions are assumed as the area of the working robot in the image.

2.2 3-D Position Detection

The 3-D position of the working robot is calculated using stereo vision theory. Even if the observing robot has only one camera, it can move to change its positions so that two or more images of the working robot are captured from different views. The corresponding points of feature are then obtained by the images from different angles. We can apply stereo vision techniques such as 8-point algorithm (Shi and Tomasi, 1994) to those corresponding points to compute the 3-D position.

3 FEATURE POINT DETECTION BASED ON SURF DESCRIPTOR

In order for fast detection of the feature points, we have tried to apply SURF (Speeded Up Robust Features) descriptor (Bay et al., 2008). SURF is improved algorithm of SIFT descriptor (D.G.Lowe, 1999) so that it can compute fast.

The SIFT descriptor is capable of robust detection of feature points in an image. It is also able to describe quantities of detected features robustly to the change of scale, illumination, and rotation of image. It is, therefore, useful for object detection and recognition. The processes of the detection of SIFT features consist of *extraction of key points*, *localization*, *computation of orientation*, and *description of quantities of features*. In the process of the *extraction of key points*, DoG (Difference of Gaussian) is used for searching local maxima to detect the positions and scales of features. Some points are then picked up from them by the process of *localization*. The orientations for those points are then computed, and their quantities of features are described. To describe the quantities of features based on the orientation, surrounding region divided by 4×4 blocks at a feature point is rotated to the direction of the orientation. Making a histogram on 8 directions for each block produces a $128(4 \times 4 \times 8)$-dimensional feature vector. The quantity of SIFT feature is represented by this vector.

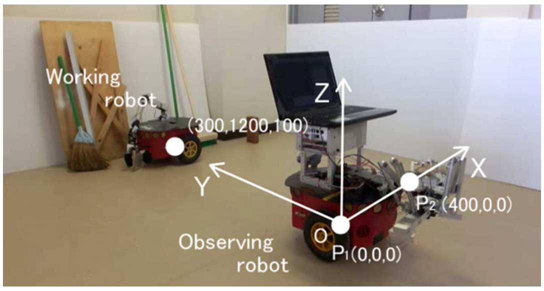

Figure 1: Experimental setup.

The SURF descriptor is improved to be faster in the above processes of *extraction of key points* and *description of quantities of features*. In the process of *extraction of key points*, SURF create the DoG image by the determinant of Hessian using a box filter instead of Gaussian function. The box filter is an approximate image of second derivative filter of Gaussian. Using the box filter, the filtering computation becomes fast because it consists of pixels which have same values so that we can obtain integral image in advance. In the process of *description of quantities of features*, the dimension of the feature vector is reduced to 64 from 128 by dividing orientation of each block into 4 directions.

4 EXPERIMENTS

4.1 Experimental Setup

The method described above has been implemented to two wheeled-mobile robots, Pioneer P3-DX (Mobile Robots Pioneer P3-DX, 2007), which is 393 mm in width, 445 mm in length, and 237 mm in height.

One robot has a camera, Canon VC-C50i, which is able to rotate in pan and tilt directions so that it is qualified as the observing robot. A board computer, Interface PCI-B02PA16W, was also mounted in order to process images from the camera in observation as well as control the movement of the robot. We utilized OpenCV for developing software for the image processing in observation.

Fig. 1 shows experimental setup. The observing robot initially stays at P_1 to observe and detect the working robot, which does not have a camera, based on the method described above. The observing robot then changes the position from P_1 to P_2 in Fig. 1 to observe the working robot in different visual angles in order to obtain its corresponding points and calculate its position.

4.2 Robot Detection

We extracted SURF features of the working robot from registered images in advance. These features were used for definition and detection of them by the observing robot. The observing robot extracted features from an input image and obtained correspondences for the working robot to detect it.

We then applied the voting method to the robot detection. Fig. 2 shows the result. Top panel shows the result when the voting method was not applied. Bottom panel shows the result when the voting method was applied. Each left panel is a registered image as the working robot. Each right panel is an image used for detection and divided into rectangle regions. The correspondences are indicated by solid brown lines. In the right-top panel, the obtained corresponding points are shown by blue circles and the extracted rectangle regions are surrounded by red line. In the right-bottom panel, the extracted points by the voting method are shown in pink circles and the rectangle regions surrounded by red line show detected regions in which more than three feature points are corresponded to those in the registered image.

The result showed the effectiveness of the voting method. When the voting method was not applied, improper regions were detected at the left bottom area in the right-top panel. On the other hand, When the voting method was applied, such improper regions were not detected and the regions which are on the robot are only detected as shown in the right-bottom panel.

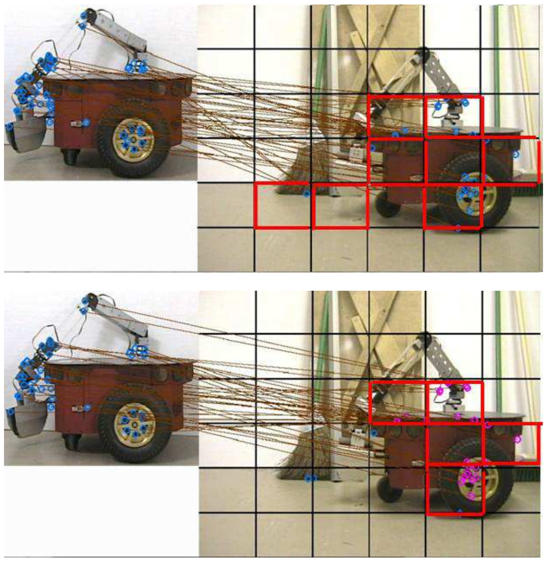

Figure 2: Experimental result of the working robot detection.

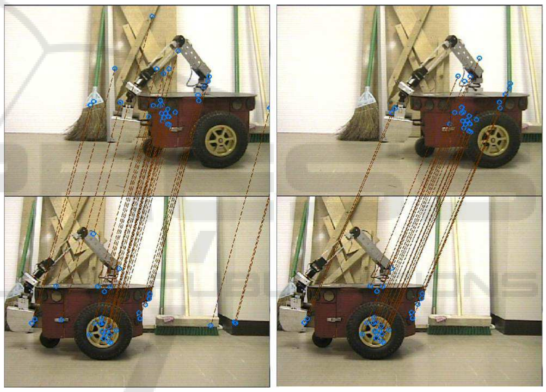

Figure 3: Correspondences between two images from P_1(top panel) and P_2(bottom panel) without voting method (left panel) and with voting method (right panel).

4.3 Position Detection

Fig. 3 shows correspondences of feature points between two images from P_1 and P_2. The top panel shows the image taken at P_1 and the bottom one shows the image taken at P_2. The left panel shows the result when the voting method was not applied. The right panel shows the result when the voting method was applied. Each corresponding points are connected by a solid line each other. In the result of the right panel, unrelated corresponding points to the robot were completely excluded. This result shows that we can obtain correct points to compute 3-D position of the working robot by using the voting method.

Table 1 shows the errors of computed 3-D positions of the working robot by the two different methods; (a) shows the result when the voting method was not applied, and (b) shows that when the voting method was applied. The errors of detected position values to actual position values on X, Y, and Z directions are described. In both results, SURF descriptor was used for detecting feature points. In the result (a), the error on Y was large. The result of (b) shows that the detection accuracy was improved specially on Y even though it was almost same on the other directions. These results showed that effective 3-D position detection is possible by the use of the voting method.

Table 2 shows the errors of computed 3-D positions of the working robot by using SURF and SIFT descriptors. The errors of detected position values to actual position values on X, Y, and Z directions are described. The processing time in each case is also described at the last row. In both results, the voting method was also applied as (b) in Table 1. The result shows that each accuracy in the use of SURF is almost same to that in the use of SIFT. With respect to the processing time, however, the use of SURF de-

Table 1: Error values on X, Y, and Z directions in 3-D positions computation of the working robot in different two ways; (a) the voting method was not applied, and (b) the voting method was applied.

	(a)	(b)
error on X	5% (+14 mm)	4% (+12 mm)
error on Y	72% (+862 mm)	34% (+404 mm)
error on Z	2% (-2 mm)	4% (+4 mm)

Table 2: Comparison of 3-D position detection of the working robot between SURF and SIFT descriptors. Error values on X, Y, and Z directions are described. In addition, processing time in execution was also shown. In both result, the voting method was applied.

	SURF	SIFT
error on X	4% (+12 mm)	8% (+24 mm)
error on Y	34% (+404 mm)	33% (+395 mm)
error on Z	4% (+4 mm)	9% (-9 mm)
time	1112 ms	1988 ms

scriptor was significantly much faster than the use of SIFT. From these results, we can see that SURF is very useful with respect to both of accuracy and execution time.

5 CONCLUSIONS

This study described fundamental detection of the positions of a working robot for assisting its object handling task based on an observation by another robot in the situation that the working robot can not perceive the object. The fundamental experiments confirmed processes in the proposed method: detection of the working robot and its 3-D position computation based on the correspondences of SURF features. Our future work will proceed with consideration of detection of the other information: position of the target object, hand motion, and so on. We will also try to expand this method to practical case toward real-time cooperation.

REFERENCES

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359.

D.G.Lowe (1999). Object recognition from local scale-invariant features. In *Proc. of IEEE International Conference on Computer Vision*, pages 1150–1157.

Mobile Robots Pioneer P3-DX (2007). In *http://www.mobilerobots.com*.

Shi, J. and Tomasi, C. (1994). Good features to track. In *Proc. IEEE Conf. on Computer Vision Pattern Recognition*, pages 593–600.