View-invariant Planar Object Detection for VisTRo

Jiyoung Jung, Yekeun Jeong, Joon-Young Lee, Hanbyul Joo, and In-So Kweon
Department of Electrical Engineering, KAIST, Republic of Korea
Electronics and Telecommunications Research Institute, Republic of Korea
{jyjung,ykjeong,jylee}@rcv.kaist.ac.kr, hbjoo@etri.re.kr, iskweon@ee.kaist.ac.kr

***Abstract*— We present a simple and powerful method for planar object detection which is robust against changes in camera positions using image warping. Applying stereo camera, we efficiently reconstruct the sparse 3D scene and find the dominant plane. We then place a virtual camera facing directly at the selected plane from the predefined distance and angle so that the plane of interest appears in the intended size without any affine distortion due to a slanted view point. The exact 3D location of any point on the surface can be recovered by oriented chamfer matching using a single template and inverse warping. This algorithm shows a good performance on indoor robot vision application which often requires detecting planar objects such as locating an elevator button or recognizing signs on the wall.**

***Keywords*— Template matching, planar object alignment.**

1. Introduction

The field of object detection and alignment has been widely explored from the perspective of robot vision application. The human visual system often perceives an object on the basis of its shape alone [3], [5]. The robot vision technology takes advantage of the shape information in the form of template matching [2], [4]. Shape-based object detection is particularly useful to the robot vision applications because the pre-required data, which is the template for matching, is as simple as Figure 1-(c) in red.

One of the major difficulties in template matching is that the object of interest may appear differently from the prepared template. In case of objects with rigid bodies, differences in appearance are mainly in scale or affine distortions caused by different camera positions. The conventional template matching cannot be free from the problems and it has to prepare multiple templates over various scales and distortions.

Therefore, what we need is the fixed angle and distance to view the object. If we recover the virtual view from the same position which the prepared template is seen from, then the object of interest must appear the same with the template. We recover the 3D shape of objects which is invariant to all the mentioned problems and provides a projective transformation to the known camera position.

2. Planar Surface Detection and Image Warping

We focus on detecting planar or near-planar objects because this assumption simplifies the complicated 3D transformation into the 2D projective homography. Given a pair of stereo images, we reconstruct 3D points of the matched edge pixels and find the most probable plane of interest. The proposed algorithm does not reconstruct dense 3D from stereo images. We detect edges from the images [1] and reconstruct 3D points from matched edge pixel pairs only. Since we do not plan to match the given template with the reconstructed 3D points directly, the sparse 3D points are sufficient because they are used to determine the dominant plane.

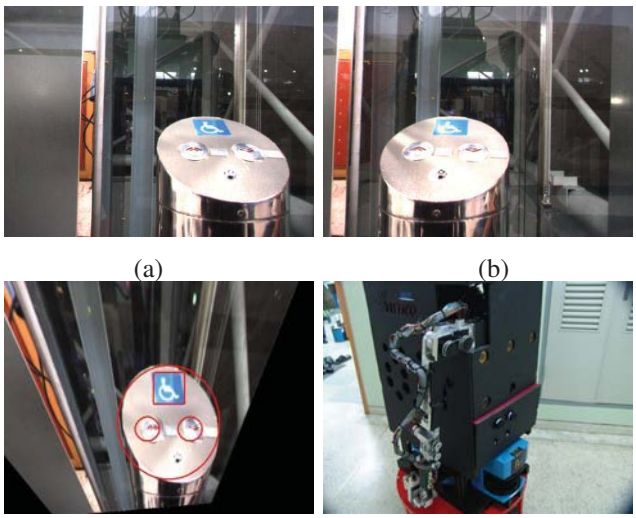

Fig. 1. A challenging example of localizing an elevator button. We extract edges from a rectified stereo image pair (a,b) and retrieve their 3D coordinates. The dominant plane is searched in 3D space and one of the input images is warped to match the template (c). The algorithm is applied to our service robot system, VisTRO (d), for the task of locating elevator button.

Once we decide the dominant plane in 3D space, we calculate the angle between its normal vector and the principal axis of the current camera. Then we place a new virtual camera in 3D space so that the image plane of the camera is parallel to the selected plane from the predefined distance. The intersection between the principal axis of the camera and the plane is set to be the center of mass of the recovered 3D points on the plane.

We then calculate homography H that warps the input image to be seen from the virtual camera. Let $N = [n_1, n_2, n_3]^T$ be the normal vector of the dominant plane, and let R and T denote the relative rotation matrix and translation vector between the current and the virtual cameras, respectively. Let d be the predefined distance from the dominant plane to the optical center of the virtual camera. The planar homography

Fig. 2. Planar or near planar objects of four categories are shown. Each pair includes the left image of a stereo input image pair and the final result, which shows image warping and oriented chamfer matching using a single template.

H is defined as follows:

$$H = R + \frac{1}{d} T N^T \quad (1)$$

The homography between the new image seen from the virtual camera and the input image instead of the dominant plane in 3D is required because we do not reconstruct dense 3D points which requires very wasteful computations. However, calculating additional homography to the virtual camera is much more efficient and likely to yield a better result than to warp the dense 3D points on the dominant plane in which each point is independently triangulated and the warped image naturally shows a very noisy result.

3. Experimental Results

We have experimented the proposed method to our indoor service robot, VisTRO, for the task of locating various indoor planar objects, as shown in Figure 2. In Figure 3, four graphs show the results of object alignment using Chamfer template matching. The dataset contains four different objects seen from four levels of distances and nine levels of angles. The results are evaluated by measuring the average pixel distance of misalignments between corresponding points on the image and the template. The proposed method using image warping and a single template shows less errors than using a single template or even multiple templates without image warping in various scale and severe affine distortions.

4. Discussion and Conclusion

We have presented a robust and efficient method for shape-based object detection and alignment. In the use of stereo camera, which is now available in most robot systems, the image is warped to be seen from the virtual camera at the predefined position relative to the plane of interest so that the object appears similarly with the prepared template. A single template is sufficient to detect objects in various sizes with affine distortions. Moreover, it shows a better performance than using multiple templates without image warping beforehand.

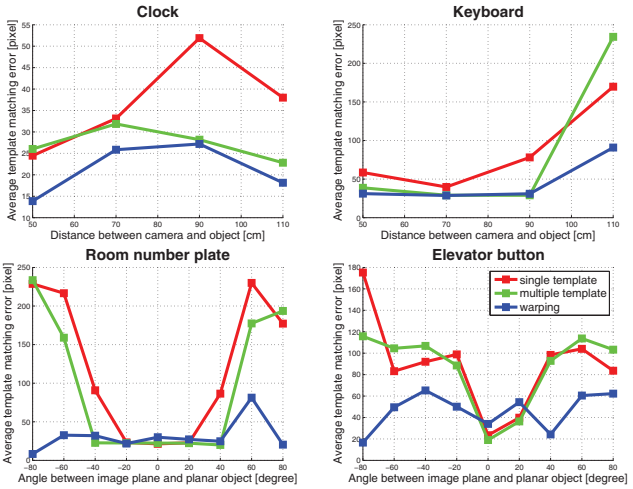

Fig. 3. Error measurements of template matching. The proposed method using image warping and a single template is shown to be more robust to scale difference and affine distortions due to different camera positions than using a single template or multiple templates without image warping.

Acknowledgements

This research was supported by MKE(Ministry of Knowledge Economy), Korea, under the Human Resources Development Program for Convergence Robot Specialists support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2010-C7000-1001-0007).

References

[1] J. F. Canny, "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679-698, 1986.
[2] T. Cour and J. Shi. "Recognizing Objects by Piecing Together the Segmentation Puzzle", IEEE Conference on Computer Vision and Pattern Recognition, 2007.
[3] H. Joo, Y. Jeong, O. Duchenne, S. Ko and I. Kweon, "Graph-Based Robust Shape Matching for Robotic Application", International Conference on Robotics and Automation, 2009.
[4] A. Opelt, A. Pinz and A. Zisserman, "A Boundary Fragment Model for Object Detection", 9th European Conference on Computer Vision, 2006.
[5] J. Shotton, A. Blake and R. Cipolla, "Contour-based Learning for Object Detection", International Conference on Computer Vision, 2005.