Image Warping for View-invariant Object Matching using Stereo Cameras

Jiyoung JUNG, Yekeun JEONG, Joon-Young LEE and In-So KWEON

Department of Electrical Engineering, KAIST
{jyjung, ykjeong, jylee}@rcv.kaist.ac.kr iskweon@ee.kaist.ac.kr

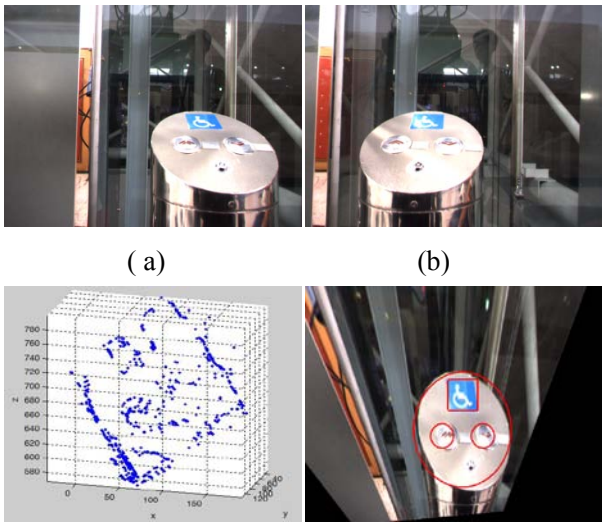
Abstract In this paper, we propose a method to increase the robustness and the efficiency of the object matching. Instead of preparing templates of various scales and deformations, we try to standardize the size and the view point of the given input image to the single template. This approach can solve the chronic problem of scale-dependence and distortion in template matching, and therefore be applied to the object detection and alignment task of a general robot system. The process has shown a good performance in the application of the service robot, Vistro, which attended the Robot Grand Challenge 2009 in Pohang, Korea.

1 Introduction

The field of object detection and alignment is widely explored from the perspective of robot vision application. Among many features, shape of an object can give abundant information about the object. Human visual system often recognize an object on the basis of the shape alone [1,2,3,4,5,6]. One popular way for object detection using shape information is to prepare a rough silhouette of the object of interest as a template and find a match within input images, which is referred to as template matching. For the robotic application, this kind of shape-based object detection is particularly useful because the template to be prepared for matching is as simple as Fig.1(d) in red.

The major problem of the template matching is that since the position of the camera can be anywhere, the object of interest in the input image may appear very differently. It may appear in various sizes and even have some distortions in shape due to different camera angles. Therefore we have to prepare several versions of templates to recognize a single class of object to handle different sizes and distortions.

The main contribution of this paper is to propose a solution to the chronic problem of scale-dependence and distortion in template matching by simply applying stereo cameras which now become a basic specification to any kind of robot system. Given a pair of stereo images, we reconstruct 3D points of the matched pixels and find the most probable plane of interest. The plane is then warped to the standardized image patch of certain size and view point. The template matching algorithm is performed by comparing



(a) (b)

(c) (d)

Fig 1: Challenging example of localizing elevator button for the service robot, Vistro. We extract edges from rectified stereo input image pair (a,b) and reconstruct 3D points. The dominant plane is searched in 3D space (c) and warped to match the template (d). The template is indicated in red.

the warped image patch and the single template we have prepared.

The remainder of this paper is organized as follows: Section 2 presents our algorithm including edge-stereo, 3D reconstruction and dominant plane search in 3D space. Image warping and template matching steps are explained in Section 3. Section 4 shows our results with real case challenging examples. A discussion and summary of this work is presented in Section 5.

2 Selecting a planar surface in 3D to warp

The first step is to determine the appropriate planar surface in 3D space to warp. The selected plane should contain the object of interest. Our assumption is that this plane of interest is not homogeneous, and therefore has a lot of edges to be detected. Sometimes we can take advantage of previous knowledge about the plane of interest such as the angle of the plane normal vector.

The proposed algorithm does not reconstruct dense 3D from stereo images. We detect edges from the images and reconstruct 3D points from matched edge pixel pairs only. Since we do not plan to match the given template with 3D points, the sparse 3D points are sufficient to determine the dominant plane. Template matching is performed on 2-dimensional image plane after warping the dominant plane.

2.1 Edge stereo

Consider a pair of rectified stereo images as input, we extract Canny edge to find the correspondences. Since the images are already rectified, the search range of correspondence is limited to horizontal direction. For each edge pixel in one image, its surrounding local window is compared with every local window surrounding each edge pixel with the same y-coordinate in the other image. The normalized cross correlation (NCC) is used for the measurement of comparing two local windows. The horizontal search range can be limited if we have previous knowledge of rough depth range for the object of interest. This kind of depth range prior can be achieved by a laser range finder installed in the robot system.

2.2 3D reconstruction

Using the projective matrices for stereo cameras and sparse correspondences for edge pixels, we can triangulate the locations of 3D points. Since z-coordinate values of the reconstructed 3D points are initially discrete due to the integer values of pixel disparity, we further refine them using 1-dimensional Kanade-Lucas-Tomasi (KLT) feature tracker.

2.3 Dominant plane search

Plane parameters can be determined by three points in 3D space. Parameters for the dominant plane in 3D space are recovered using RANSAC algorithm. Among the planes parameterized by three randomly selected points, we consider the dominant plane has the most inliers. We may have some previous knowledge about the dominant plane such as the angle of the plane normal. The dominant plane should satisfy such constraints while having the most 3D points on the plane.

3 Warping

Once we decide which plane to warp, we calculate the angle between the dominant plane normal vector and the standardized plane normal vector. The standardized plane is a virtual plane in 3D space which is parallel to the image plane. Among the recovered 3D points on the dominant plane, the point of mean x and y coordinate should be on the principal line of the camera.

We then calculate the rotation and translation between the dominant plane normal vector and the virtual standardized plane normal vector. Let $\mathrm{N} = [\mathrm{n}_1, \mathrm{n}_2, \mathrm{n}_3]^{\mathrm{T}}$ be the dominant plane normal vector, and let R and T denote the rotation and translation matrices respectively. Let d>0 denote the distance from the virtual standardized plane to the optical center of the camera. The planar homography H is defined as follows:

$$\mathrm{H} = \mathrm{R} + \frac{1}{\mathrm{d}}\mathrm{T}\mathrm{N}^{\mathrm{T}} \quad \in \mathbb{R}^3$$

Now we have obtained the homography between the dominant plane in 3D space and the virtual standardized plane. In order to warp the image from one of the input image pair to the virtual plane, we have to

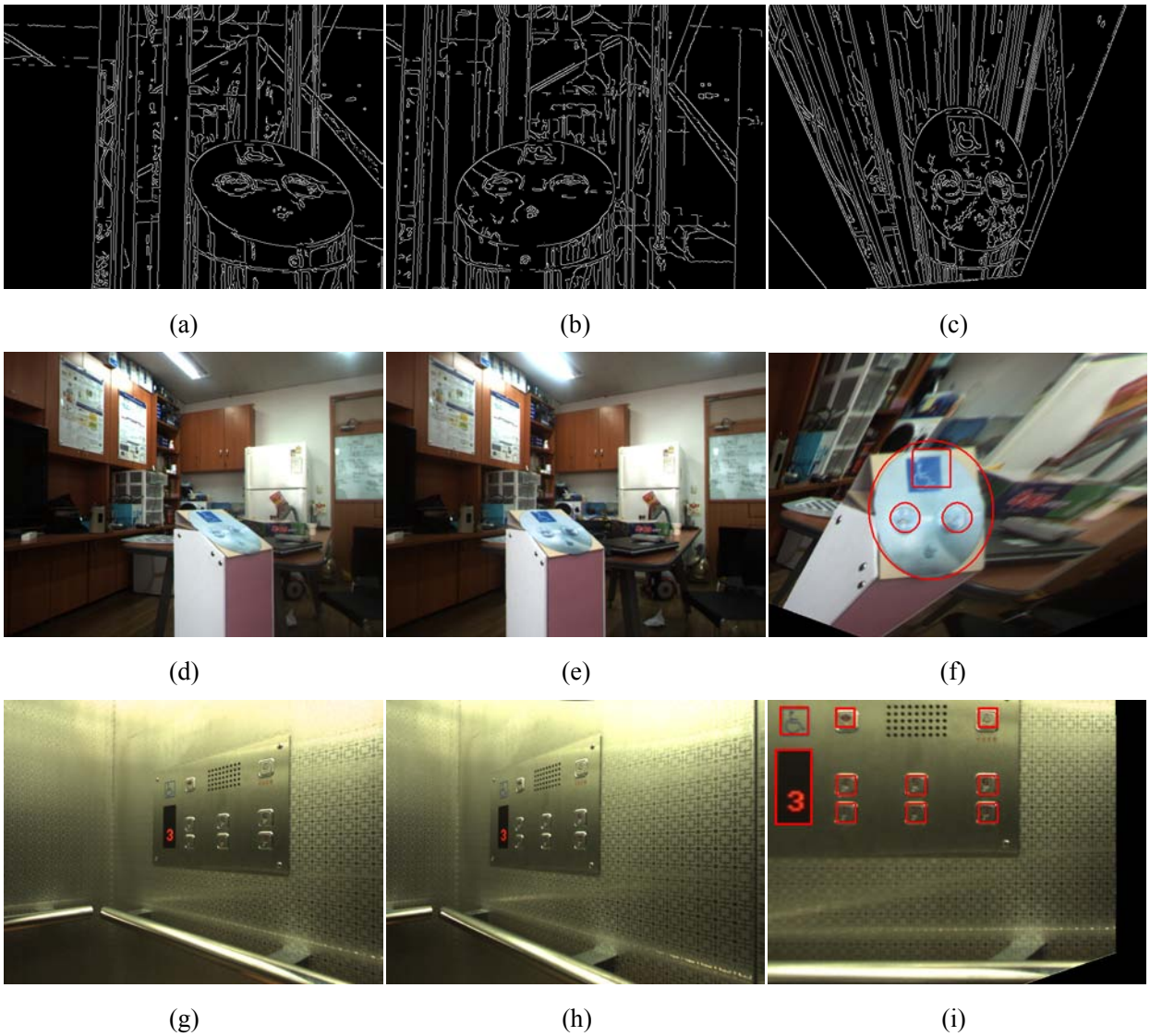


Fig 2: Figures in the top row show the extracted edges from figure 1-(a,b,d) which are the major feature we used in this work. The popularly used Canny edge feature is good enough for our sparse 3D reconstruction. Figures (d-i) show other experimental results. Figures (d,e,g,h) are the demonstration of severe distortion due to side view point. Images are warped to have the same view point as the prepared templates and then matched (f,i).

find the homography between the input image and the virtual plane. This additional homography is required because we did not reconstruct dense 3D points. However, calculating additional homography to the virtual plane is much simpler and likely to yield a better result than to warp the dense 3D points on the dominant plane.

4 Experiments

In order to show the effect of our proposed method, we performed chamfer matching. Chamfer matching is used to correlate model template with edge image. It is robust in clutter and illumination, but it is difficult to apply in real images because the chamfer distance is vulnerable to scale, rotation and translation. Therefore, multiple templates are needed to handle those geometric variations.

In the proposed method, geometric variations like scale and rotation are dealt with standardization of the given input image. Therefore, we can apply chamfer matching using a single template.

Figure 2 shows the overall process of our proposed method.

5 Discussion and conclusion

We have presented a robust and efficient method for shape-based object detection and alignment. In the use of stereo cameras, which is now available in most robot systems, the image is warped to the virtual plane of specific scale and viewpoint. This allows us to prepare only one template to match the standardized input image. This can be a simple but useful solution to the problem of scale-dependence and deformation in template matching.

References

[1] A. Opelt, A. Pinz, and A. Zisserman: A boundary fragment model for object detection, ECCV, vol. 2, pp. 575–588, 2006.

[2] J. Shotton, A. Blake, and R. Cipolla: Contour-based learning for object detection, ICCV, 2005.

[3] A. Thayananthan, B. Stenger, P. Torr, and R. Cipolla: Shape context and chamfer matching in cluttered scenes, CVPR, pp. 127-134, 2003.

[4] D. M. Gavrila: Pedestrian detection from a moving vehicle, 6th European Conf. on Computer Vision, vol. 2, pp. 37–49, 6 2000.

[5] T. Cour and J. Shi: Recognizing objects by piecing together the segmentation puzzle, CVPR, 2008.

[6] H. Joo, Y. Jeong, O. Duchenne, S. Ko, and I. Kweon: Graph-Based Robust Shape Matching for Robotic Application, ICRA, 2009.