

PAPER • OPEN ACCESS

Transparent object detection and location based on RGB-D camera

To cite this article: Chen Guo-Hua *et al* 2019 *J. Phys.: Conf. Ser.* **1183** 012011

View the [article online](#) for updates and enhancements.

You may also like

- [Consistent-Resolution Network for 3D Hand Shape Estimation from a Single RGB Image](#)
Qi Wu, Joya Chen, Zhiming Yao et al.
- [Comparison of single-spot technique and RGB imaging for erythema index estimation](#)
I Saknite, A Zavorins, D Jakovels et al.
- [Identification of cavitation intensity for high-speed aviation hydraulic pumps using 2D convolutional neural networks with an input of RGB-based vibration data](#)
Qun Chao, Jianfeng Tao, Xiaoliang Wei et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Transparent object detection and location based on RGB-D camera

Chen Guo-Hua, Wang Jun-Yi and Zhang Ai-Jun

Beijing University of Chemical Technology

E-mail: chengh@mail.buct.edu.cn

Abstract. In order to improve the accuracy and efficiency of robot grasping, we propose a new method for transparent object detection and location that utilize depth image, RGB image and IR image. In detection process, an active depth sensor (RealSense) is firstly employed to retrieve the transparent candidates from the depth image and the corresponding candidates in the RGB image and IR image are then extracted separately. A transparent candidate classification algorithm is subsequently presented that uses SIFT features to recognize the transparent ones from the candidates. In location process, we obtain a new group of RGB images and IR images by adjusting camera orientation to make its optical axis perpendicular to the normal direction of the plane on which the object is placed. The object contours in RGB image and IR image are then extracted, respectively. The three-dimensional object is finally reconstructed by means of stereo matching of the two contours, and the current pose information of the object is calculated in the end. In order to verify the feasibility of the method, we built a hand-eye test system with a movable industrial robot to detect and capture transparent objects at different locations. The final test results demonstrate that the method is more general and effective than the traditional one.

Keywords: Transparent Object Detection, location, RGB-D Camera, Depth Image, Robotic Grasping.

1. Introduction

Transparent objects are very common in our daily life, such as glasses and bottles in the home, beakers and test tubes in the laboratory. With the development of robotics, robots have been widely used in many fields, such as service robots in the domestic environment and industrial robots in the laboratory environment. These robots inevitably involve the identification and grasping of transparent objects. Particularly in unmanned laboratories, transparent object identification and grasping is especially a key issue.

Object recognition has been a hot research topic in computer vision, but transparent object recognition is a difficult problem to solve, because transparent materials are difficult to detect due to the appearances of transparent objects that can change over different backgrounds. We use the advantage of the Kinect sensor, which provides scene depth information using a projected pseudo-random IR pattern, but it cannot form effective depth data on the transparent object surface.

In order to solve this problem, we present the detection and pose recognition of transparent objects based on RealSense's RGB-D information and movable robotic arm and we propose several methods in this paper. Firstly, describe the previous work has been done. Secondly, we propose algorithm to segmentation and recognition. Next, we propose a new method to reconstruction the objects. Then, we



use the model to pose estimation of rigid transparent objects. Finally, we evaluate the proposed approach by using this method in five objects and investigate performance in various scenes containing transparent objects.

2. Related work

There are three types of recognition method of transparent objects: 2D Computer Vision, 3D Computer Vision and others.

2D Computer Vision. The appearance of a transparent object strongly depends on lighting and its background. Transparent objects usually don't have their own texture features, their edges are typically weak and intensity gradient features are heavily influenced by seeing through background clutter as well as secularity edges induced by lighting. So that classical computer vision algorithms for recognition and pose estimation are difficult to apply to transparent objects. To solve these problems, an Additive Latent Feature Model is used for Transparent Object Recognition as [1] showed, they used a LDA-SIFT model, [2] proposed a new feature called the light field distortion (LFD) feature. Based on [2], [3] showed a LFD feature and use it to segment the transparent objects from RGB image, and it got a better result. [4] showed a video of a transparent object shot under varying illumination to estimate the normal map, [5] combined the systematic distortions in background texture occurring at the boundaries of transparent objects with the strong lightened typical of glass surfaces to train a hierarchy of classifiers, identify glass edges, and to find consistent support regions for these edges. [6] employed a probabilistic formulation for segmentations glass regions, similar to the [5], [7] proposed to use binary criteria to guide the segmentation process, [8] reconstructed a transparent surface modelling from a pair of polarization images.

3D Computer Vision. 3D point clouds are successfully used for object recognition and pose estimation. However, modern sensors (Kinect, ToF cameras, stereo cameras, laser scanners) can't estimate depth reliably and produce point clouds for transparent and specular objects so these algorithms cannot be applied. Cross-modal stereo can be used to get depth estimation on transparent objects with Kinect as in [14], but its quality is inferior. Acquisition of 3D data and reconstruction of transparent objects is a challenging and unsolved problem. [14] showed the method of using missing depth information to segment the object and pose estimate which first proposed by [19], [16] used Geometric Hashing to solve the clutter. [12] reconstructed objects by moving around a transparent object in the scene, [13] combined multiple sensor modalities, [15] showed a method of using seashell sensors to add missing point cloud data. [19] improved the Cross-modal stereo by using fully-connected CRF.

There also have some other methods by using non-image sensors, such as using colour image and laser reflectance image that extracting object via reflectance laser intensity, or using a stereo camera and three-dimensional laser radar system to simultaneously illuminate transparent objects by a plurality of light sources and identify transparent objects by the high spot.

But, the quality of transparent object recognition by using traditional RGB information is highly dependence on the transparency of the object and the ambient light, and these methods also have certain limits of shooting angle of an object and cannot obtain the depth information form the object. And the method that uses non-image sensor is very complex that cannot suitable for the needs of daily life. The traditional use of stereo visual recognition needs to advance the identification of objects such as modelling operations. So the traditional identification process is very complex.

3. Recognition of Transparent Objects

RealSense uses the laser speckle pattern technique to estimate the depth, it can project infrared light onto the surface of the object. However, it cannot form an effective pattern on the surface of transparent objects and also cannot get the effective point cloud information. Based on this condition, we can use the area where has no depth information as the region of interesting in depth image, and segment this region in the RGB image. Then we use the SIFT feature to detect the change of the feature points of the segmented region. According to the movement of feature points to determine whether the object is transparent.

3.1. Transparent Candidate Retrieval

The method we used is based on [11], but we improved the method to exclude non-transparent objects which can generate invalid depth areas. RealSense not only cannot estimate the depth of transparent objects, and it also cannot estimate contours of objects. Sometimes it can receive the depth information of the background behind transparent objects as figure 1, but we can take advantage of this fact: regions where RealSense cannot form effective depth information are likely to belong to transparent objects. So this cue can be used to segment transparent objects on an image. We can use this cue to generate the mask of transparent objects. And we use GrabCut segmentation algorithm on RGB image, we can get the contours and area of the object as figure 2.

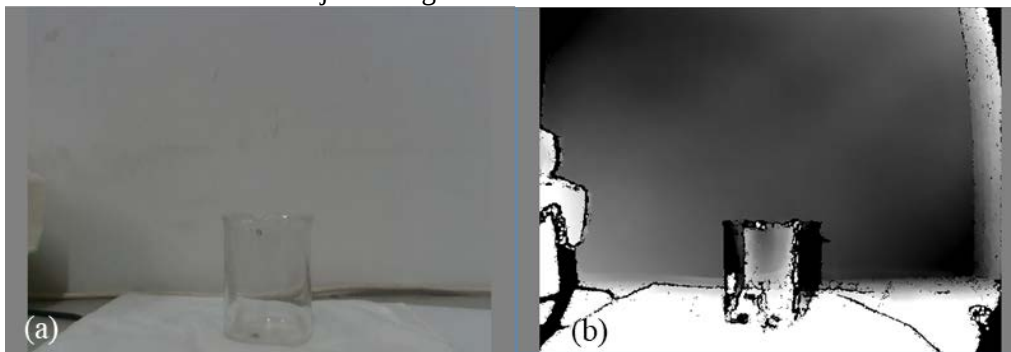


Figure 1. (a) RGB image (b) Depth image

Before we get the interesting region of the depth image, firstly we use the formulas as (1) and (2) to the depth image to remove the noise caused by the infrared beating of the device itself.

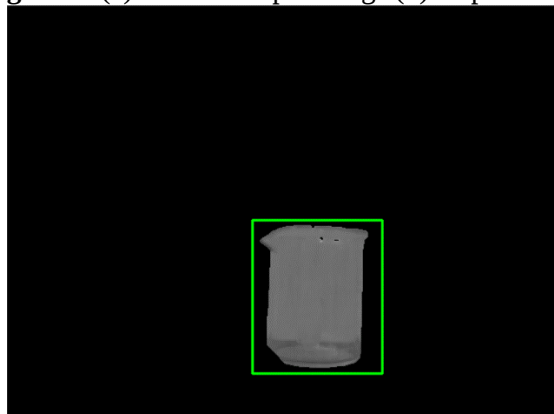
$$w = w_g \times w_c \quad (1)$$

$$w_c = \exp\left(-\frac{(G(i,j)-G(x,y))^2}{2\sigma_c^2}\right) \quad (2)$$

Where (i, j) is the standard deviation of the Gaussian function, and G (i, j) is the standard value when the depth image is transformed into a grayscale image at the pixel (i, j)) at the grey value.

After filtering, morphological operations (closing and opening) are applied to find and eliminate small regions in the depth image, the result is a mask of transparent objects (figure 2(a)-2(b)). Using the mask information, the Grab Cut segmentation algorithm is used to segment the area of the transparent object on the corresponding RGB image and IR image. We can use the mask information as the initial conditions and constraints of the algorithm when we use Grab Cut, we can get the final segmentation results that are shown in figure 3.



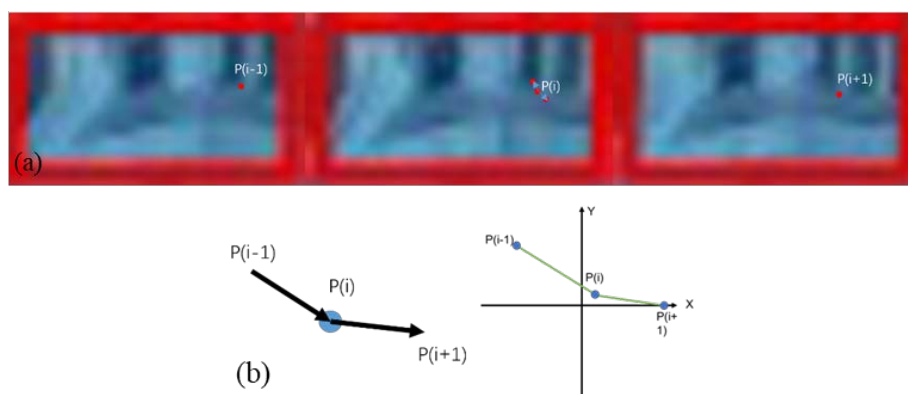
Figure 2. (a) Mask of depth image (b) Depth image**Figure 3.** Results of glass segmentation

3.2. Transparent Candidate Classification

Transparent objects are made of refractive materials; such as glass or plastics. It can distort rays emanating from the scene background (figure 4). We use the phenomenon of background distortion caused by refraction to identify transparent objects.

**Figure 4.** Background distortion from changing viewpoints

We use the SIFT feature to classification the transparent candidate. As described in figure 5, due to refraction of the transparent object, background texture changes irregularly. Thus, the SIFT features obtain from the transparent object are more irregular than features from the opaque object, and these features deviate from the hyperplane due to the Lambertian reflection in the phase space. We could use optical flow method to detect the feature points. When we move the camera linearly, if the feature points move irregularly as figure 5, the area which the key points in is the transparent area.

**Figure 5.** (a) The feature points extracted from picture (b) The feature points move irregularly

We take an arbitrary feature point in the image from one of the viewpoints as $p_s(0)$, and the next few frames of this feature point as $p_s(i)$. And we use the following formulae to express the change of this viewpoint,

$$\Delta p_s(i) = p_s(i+1) - p_s(i) \quad (3)$$

$$\sigma_s^2 = \sum (\Delta p_s - \widehat{\Delta p_s})^2 / M \quad (4)$$

$$K = \sum (\sigma_s^2 - \widehat{\sigma_s^2})^2 / N \quad (5)$$

As shown in these formulae, K is the threshold. We can adjust the value of K to determine whether the region of interest is transparent, M is the number of frames that can continuously stabilize the feature points, N is the number of feature points that can continuously and steadily obtain.

4. Reconstruction

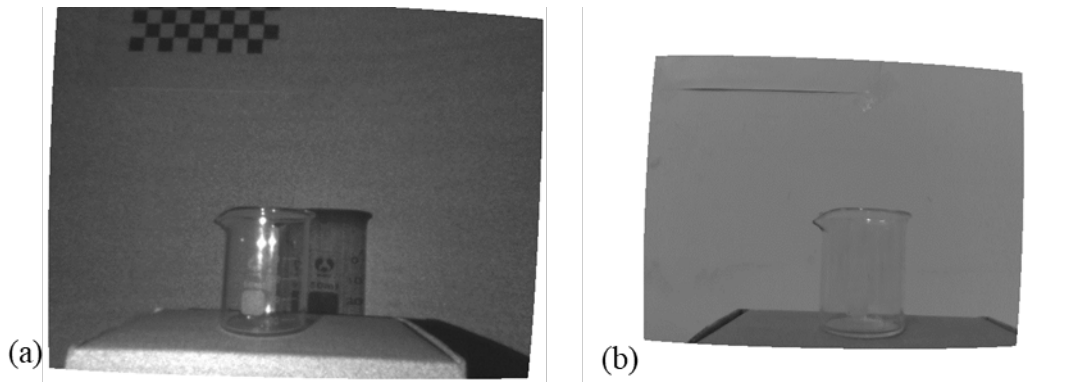
In the modelling phase, we first adjust the camera's angle, the camera axis and the object plane perpendicular to the use of cross-mode stereo matching, in multiple locations to obtain RGB images and IR images, extract the image of the object contour information on the object reconstruction. Finally, the reconstructed model is used to estimate the pose of the object.

4.1. Contour extraction

Firstly, we adjust the camera's axis position perpendicular to the normal direction of the plane where the object is located, and then we use the method mentioned earlier to segment the object to divide the contours from the RGB image and the IR image. As shown in figure 6, the object contour edge is very rough. In order to ensure that the reconstruction results are accurate, we smooth the contour information. Since the Gaussian distribution has good characteristics, and the Gaussian convolution can suppress the size of the noise by controlling the size of σ . so we use the Gaussian filter for filtering contour.

$$G(t, \sigma) = k\sigma^2 \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (6)$$

In this formula, K is a constant, we can adjust the value of k to change the filter coefficients. The filter window typically is $6\sigma \sim 8\sigma$.



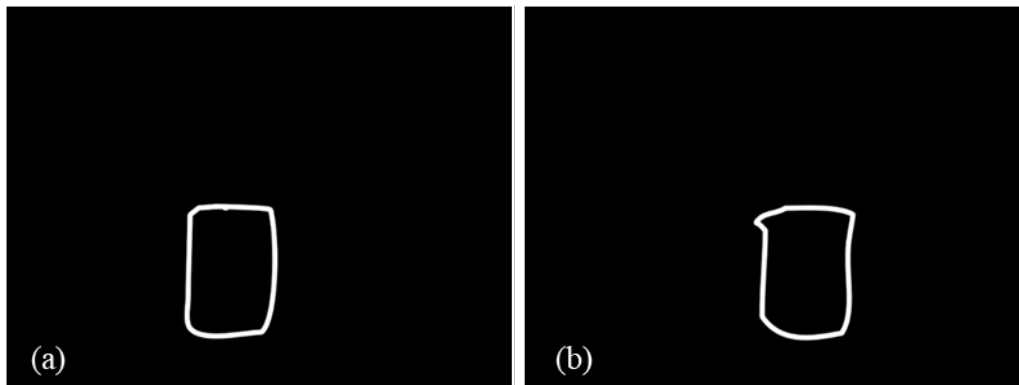


Figure 6. (a) IR image (b) RGB image (c) Contours of IR image (d) Contours of RGB image.

After we get the contours obtained from the two views, we need to fit the contours. Since the contours in the IR image are more flawed than the RGB images, so we can merge these two contours together and get the maximum of the contours. Firstly, we obtain the central moment of the two contours, then merge the central moments, finally we extract the outer parts of the two contours. As shown in figure 7.

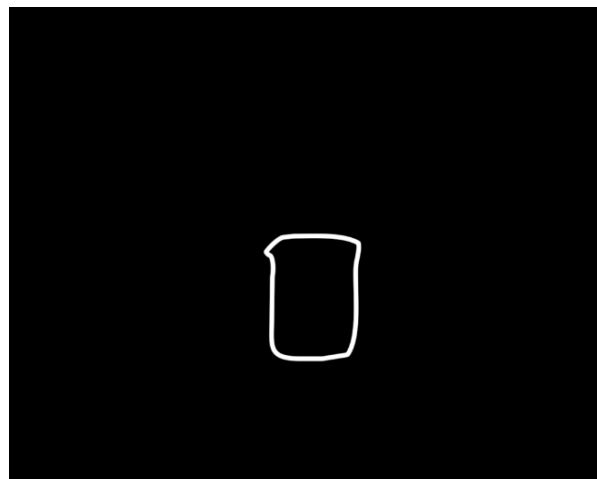


Figure 7. Merge of contours

4.2. Reconstruction

While moving the device, we extract contours from IR images and RGB images to use stereo matching to obtain the approximate mean plane information and the axis position information of the object as a three-dimensional reconstruction of the object geometric information. And we can get the object contour normalized by rotating, and get the three-dimensional information of the object. As shown in figure 8.

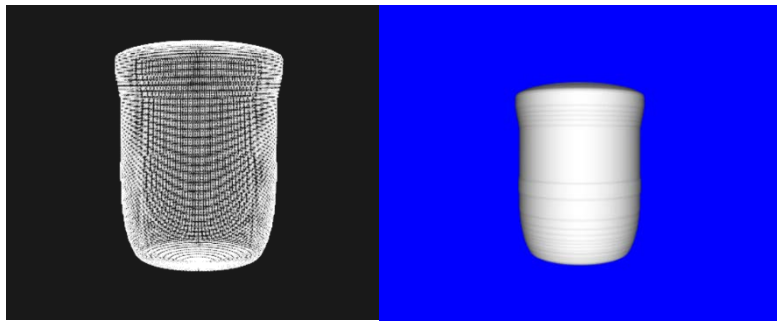


Figure 8. Point cloud data and surface reconstruction

4.3. Objects Capture

After these previous steps, we already get the object of the three-dimensional model information and the current location of the object. Because we need to capture transparent objects in the Lab finally, so we need a method to control the robot arm to capture the object accurately.

In order to make the robot grasp the object accurately, we refer to many papers of robot grasping such assume of the object's current position information, via the outline of the object to find the appropriate two points to capture and carry out effective capture behaviour. The complete process is shown in figure 9.

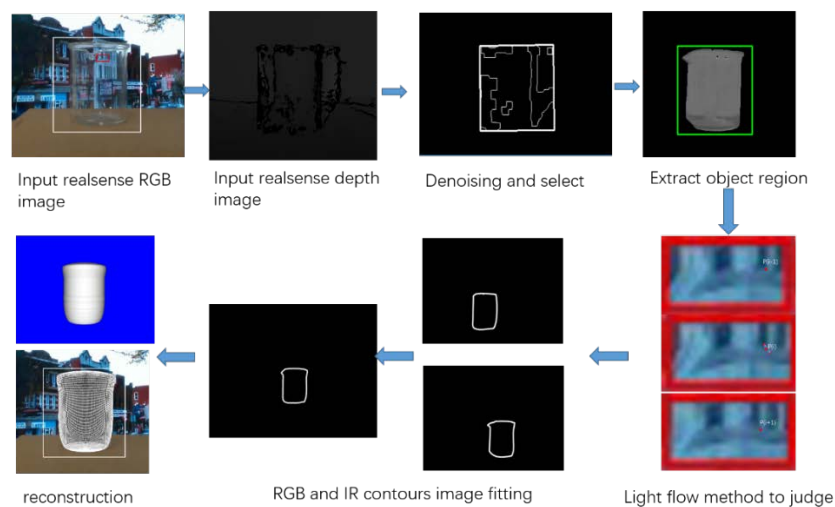


Figure 9. Process of reconstruction and pose estimation

5. Experiment

Here we use RealSense sensor to get the image, this sensor has a colour sensor (1080P@30FPS) and a depth sensor (640X480@60FPS). We fix the camera at the robotic arm to test this method.

5.1. Assumption

We evaluated our proposed method by capture the transparent objects using RV-4F robot in a laboratory setting and real environments under following assumptions;

- 1 the object is not far from the robot in order to do this task without moving the base.
- 2 that there is no mutual occlusion between transparent objects
- 3 the poses of transparent objects are fixed after the experiment begins.
- 4 the background of the object is grained.

We performed some experiments in a laboratory and real setting to evaluate robustness and limitations of our proposed method

5.2. Recognition of transparent objects

We test the performance of transparent object recognition and compare the results with the method proposed in [6]. In this test, we use accuracy and miss to evaluate the performance of transparent object recognition. Accuracy is the percentage of the number of correctly recognized transparent object over the number of all tested transparent objects. Miss is the ratio of object that is opaque, doesn't have depth information and misrecognized as an opaque one.

Miss refers to the proportion of the opaque objects that are mistaken as transparent ones in all the opaque objects. The main reason why opaque objects may be considered as transparent ones is that some opaque objects are specular objects. The ambient light, therefore, may be strongly reflected by the smooth surface during the detection process. The reflected environmental light information is regarded as the background information of the occlusion area of the object, and thus the object is consequently judged to be a transparent object.

We randomly put five transparent objects (glass goblet, beaker, test tube, cups and graduate) and other non-transparent objects that will cause unknown depth value in the scene to test the miss. In the test, we will put transparent objects and non-transparent objects in the same scene. The results of our tests are shown in table 1. As can be seen in table 1, both methods have high recognition rates of the transparent object area, but our method is better when identifying non-transparent objects. Because our method can recognize more accurate than traditional one in non-transparent objects recognition which have unknown value in the depth image.

Table 1. Accuracy and miss of transparent object recognition

Method	Accuracy	Miss
Our method	87.64%	37.93%
Method in [11]	87.64%	87.24%

5.3. Object Reconstruction

Finally, in figure 10 we show the result of our approach to reconstruct 3D models. In order to test the accuracy of our pose detection, we use the robot to capture the objects. We successfully capture 70 out of 85.



Figure 10. Left: Real scene. Right: Reconstructed model

5.4. Distance Analysis

Since the depth sensor has a range of work, so we move the camera in a difference of $\pm 15\text{cm}$ from the reference position 60cm. Figure 11 shows that the capture ratios are decreased when the displacement further from the reference position, because the accuracy of depth information of objects that sensors detect varies due to the distance.

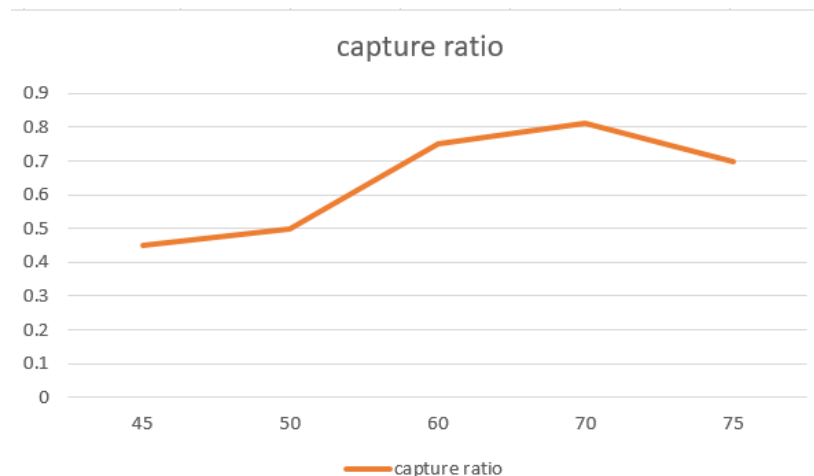


Figure 11. The relationship between Distance and grab success rate.

6. Conclusion

Using this method, the rotating transparent objects can be effectively detected and captured, and the 3D models of the transparent objects are built meanwhile. These 3D models will be saved into a database and used as source materials in the future recognition process, for example, they can be used as templates to deal with the occlusion situation in the geometric Hexa method as mentioned before [11]. Additionally, this method can be applied for autonomous mobile robots to avoid transparent obstacles. The shapes of identified transparent objects of this method are limited, for it can only deal with the rotating objects, hence in next step we will focus on expanding the identifying scope of shapes.

References

- [1] Fritz M, Bradski G, Karayev S, Darrell T and Black M J 2009 An additive latent feature model for transparent object recognition *Proc. DBLP Advances in Neural Information Processing Systems (Vancouver, Canada, 7–10 December 2009) (Red Hook, NY: Curran)* pp 558–566.
- [2] Maeno K, Nagahara H, Shimada A and Taniguchi R I 2013 Light field distortion feature for transparent object recognition. *Proc. IEEE Int. Conf. On Computer Vision and Pattern Recognition (Portland, OR, 23–28 June 2013) (New York: IEEE)* pp 2786–2793.
- [3] Xu Y, Nagahara H, Shimada A and Taniguchi R I 2015 TransCut: Transparent object segmentation from a light-field image. *Proc. IEEE Int. Conf. On Computer Vision (Santiago, Chile, 7–13 December 2015) (New York: IEEE)* pp 3442–50.
- [4] Yeung S K, Wu T P, Tang C K, Chan T F and Osher S J 2015 Normal estimation of a transparent object using a video *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 890–7.
- [5] McHenry K, Ponce J and Forsyth D 2005 Finding glass. *Proc. IEEE Int. Conf. On Computer Vision and Pattern Recognition (San Diego, California, 20–26 June 2005) (vol 2) (New York: IEEE)* pp 973–9.
- [6] Torres-Gómez A and Mayol-Cuevas W 2014 Recognition and reconstruction of transparent objects for augmented reality. *Proc. IEEE Int. Sympos. On Mixed and Augmented Reality (ISMAR) (Munich, Germany, 10–12 September 2014) (New York: IEEE)* pp 129–34.
- [7] McHenry K and Ponce J 2006 A geodesic active contour framework for finding glass. *Proc. IEEE Int. Conf. On Computer Vision and Pattern Recognition (New York, NY, 17–22 June 2006) vol1*

- (New York: IEEE) pp 1038–44.
- [8] Miyazaki D, Kagesawa M and Ikeuchi K 2004 Transparent surface modeling from a pair of polarization images *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 73–82.
- [9] Klank U, Carton D and Beetz M 2011 Transparent object detection and reconstruction on a mobile platform. *Proc. IEEE Int. Conf. On Robotics and Automation (ICRA)* (Shanghai, China, 9–13 May 2011) (New York: IEEE) pp 5971–78.
- [10] Eren G, Aubreton O, Meriaudeau F, Secades L S, Fofi D, Naskali A T, Truchetet F and Ercil A 2009 Scanning from heating: 3D shape estimation of transparent objects from local surface heating *Opt. Express* **17** 11457–68.
- [11] Lysenkov I, Eruhimov V and Bradski G 2013 Recognition and pose estimation of rigid transparent objects with a kinect sensor *Proc. Robot. Sci. Syst.* **8** 273–5.
- [12] Alt N, Rives P and Steinbach E 2013 Reconstruction of transparent objects in unstructured scenes with a depth camera. *Proc. IEEE Int. Conf. On Image Processing (ICIP)* (Melbourne, Australia, 15–18 September 2013) (New York: IEEE) pp 4131–35.
- [13] Hagg A, Hegger F and Plöger P 2016 On recognizing transparent objects in domestic environments using fusion of multiple sensor modalities *Preprint arXiv/1606.01001*.
- [14] Chiu, W W C, Blanke U and Fritz M 2011 Improving the kinect by cross-modal stereo. *Proc. Int. Conf. On British Machine Vision (Dundee, Scotland, 29 August–2 September 2011)* (Manchester, UK: BMVC Press) p 3.
- [15] Huang K, Jiang L T, Smith J R and Chizeck H J 2015 Sensor-aided teleoperated grasping of transparent objects. *Proc. IEEE Int. Conf. On Robotics and Automation (ICRA)* (Seattle, WA, 26–30 May 2015) (New York: IEEE) pp 4953–9.
- [16] Lysenkov I and Rabaud V 2013 Pose estimation of rigid transparent objects in transparent clutter. *Proc. IEEE Int. Conf. On Robotics and Automation (ICRA)* (Karlsruhe, Germany, 6–10 May 2013) (New York: IEEE) pp 162–9.
- [17] Lei Z, Ohno K, Tsubota M, Takeuchi E and Tadokoro S 2011 Transparent object detection using colour image and laser reflectance image for mobile manipulator. *Proc. IEEE Int. Conf. On Robotics and Biomimetics (ROBIO)* (Karon Beach, Thailand, 7–11 December 2011) (New York: IEEE) pp 1–7.
- [18] Saygili G, van der Maaten L and Hendriks E A 2014 Hybrid kinect depth map refinement for transparent objects. *Proc. IEEE Int. Conf. On Pattern Recognition (Stockholm, Sweden, 24–28 August 2014)* (New York: IEEE) pp 2751–6.
- [19] Ulrich M, Wiedemann C and Steger C 2009 CAD-based recognition of 3D objects in monocular images. *Proc. IEEE Int. Conf. On Robotics and Automation (ICRA)* (Kobe, Japan, 12–17 May 2009) (New York: IEEE) pp 1191–8.
- [20] Lin Y C, Wei S T and Fu L C 2014 Grasping unknown objects using depth gradient feature with eye-in-hand RGB-D sensor. *Proc. IEEE Int. Conf. On Automation Science and Engineering (CASE)* (New Taipei, Taiwan, 18–22 August 2014) (New York: IEEE) pp 1258–63.