

RGB-D Based Object Segmentation in Severe Color Degraded Environment

Chao Wang, Sheng Liu^(✉), Jianhua Zhang, Yuan Feng, and Shengyong Chen

College of Computer Science and Technology, Zhejiang University of Technology,
Hangzhou 310023, Zhejiang, China
edliu@zjut.edu.cn

Abstract. In the robotic waste sorting lines, most existing segmentation methods will fail due to irregular shapes and color degradation of solid waste. Especially in the cases of adhesion and occlusion, errors may frequently occur while labeling the ambiguous regions between solid waste objects. In this paper, we propose an efficient RGB-D based segmentation method for construction waste segmentation in harsh environment. First, an efficient background modeling strategy is designed to separate the solid waste regions from the cluttered background. Second, we propose an ambiguous regions extracting method to deal with the problem of adhesion and occlusion. Finally, a relabeling method is developed for ambiguous regions and a high precision segmentation will be obtained. A dataset of construction materials consists of RGB-D images is built to evaluate the proposed method. Results show that our approach outperforms other state-of-the-art methods in harsh environment.

Keywords: Solid waste recycling · Object segmentation
Ambiguous region extraction

1 Introduction

Construction waste recycling is most sustainable way to manage waste materials generated during construction and remodeling. Robotic waste sorting is one of the key technology in construction waste recycling. And, solid waste objects segmentation is the core technology in robotic waste sorting. Traditional treatment of construction waste is landfill, which will cause serious air and soil contamination. With the rapidly increasing amount, situation will be increasingly severe. So there is growing interest in studying the construction waste recycling. The appearance of robots provides a new, more efficient solution to this problem as they can grasp object quickly and work continuously. Image segmentation algorithms, especially with depth information, are indispensable in robotic waste sorting as they can offer solid waste objects' information of positions and contours [10–12, 15].

Custom image segmentation algorithms are not applicable in industrial field. Most 2D image segmentation methods [2, 3] use features such as color and contour. Though they achieve a good performance on some datasets, they cannot

deal with complex industrial scene. An example image of construction waste scene is shown in Fig. 1 (a). The surface of the conveyor belt is covered with dust and solid waste objects are not all isolated. Because of the color degradation caused by deposition of dust particles on the surface of the solid waste, custom 2D image segmentation methods lose their effect. With the advent of depth sensors, image segmentations with the depth information became a research focus. However, recent work on RGB-D image process has been targeted to semantic segmentation or labeling [1, 4, 16]. These methods aimed to assign a category label to each pixel of an image and complete segmentation and classification of whole scene. As for the task of solid waste sorting, objects contours and positions are indispensable to find best grab points and angles. Most semantic segmentation methods lack the concept of object instance, they can not separate adhesive objects with same category label. Furthermore, deep learning is widely used in semantic labeling such as [5, 13]. However, solid waste objects are volatile in shape and color degraded so that it is difficult to collect a representative training set. It is needed to develop a new algorithm for solid waste objects segmentation in harsh environment.

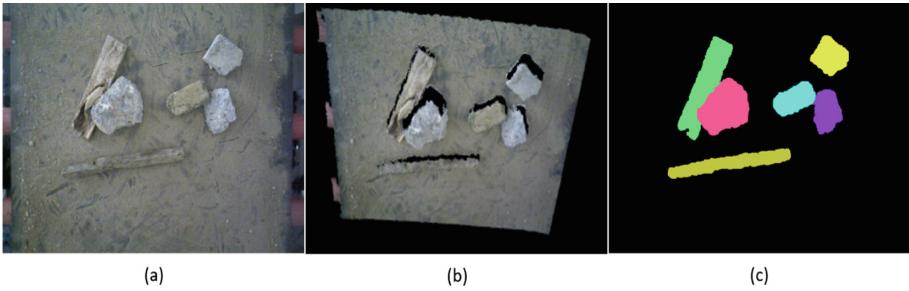


Fig. 1. A segmentation example of proposed algorithm. (a) Original image of cluttered belt scene. (b) Point cloud of cluttered belt scene. (c) Result of proposed algorithm.

In this paper, we propose an efficient RGB-D based segmentation method for construction waste segmentation in harsh environment. Our contributions in this paper are summarized as follows: (1) we propose a strategy to extract ambiguous regions to separate adhesive and occluded objects and perform pixel-level relabeling on ambiguous regions to get a high precision result. (2) We also build a unique challenging dataset of construction waste.

The paper is structured as follows. The next section discusses the presented work in RGB-D image segmentations. Sect. 3 shows the structure of the proposed method, and explains more detailed about every module. Evaluation and results are shown in Sect. 4, before the work ends with a conclusion in Sect. 5.

2 Related Work

In 3D scene, a single object consists of several planes. Segmenting an image in separate objects means labeling each plane. The approach proposed by Holz et al. [6] compute local surface normal and cluster the pixels in different planes in normal space. And the result planes are segmented and classified in different objects in both normal space and spherical coordinates. Analogously, [16] propose a multiscale-voxel algorithm to do plane extraction. Then the result planes are combined with depth data and color data to apply graph-based image segmentation. In [12], Richtsfeld et al. preprocess input RGB-D image based on surface normal, get surface patches by using a mixture of planes and NURBS (non-uniform rational B-splines) and find the best representation for the given data by model selection [8]. Then, they construct a graph from surface patches and relations between pairs of plane patches and perform graph cut to arrive at object hypotheses segmented from the scene. Irregular shapes of solid waste objects will lead to over-segmentation by plane-based methods. Moreover, plane-based methods do not deal well with touched objects caused by adhesion and occlusion.

An object is also defined as a compact region enclosed by a closed edge which is called object contour. Segmenting objects means finding effective object contours in image. In [10], Mishra et al. build a probabilistic edge map obtained by color and depth cues, and then select only closed contours that correspond to objects in the scene while reject duplicate and non-object closed contours by the fixation-based segmentation framework [9]. Toscana and Rosa [15] use a modified canny edge detector to extract robust edges by using depth information and two simple cost functions are proposed for combining color and depth cues to build an undirected graph, which is partitioned using the concept of internal and external differences between graph regions. As color degradation in construction waste images is severe, algorithms based on color cues lose their effect.

3 Our Algorithm

3.1 System Overview

Our algorithm consists of three major parts: Background Modeling, Ambiguous Regions Extraction and Ambiguous Regions Relabeling. Figure 2 shows the processing chain of those parts in detail. Background modeling builds a background model based on depth information, and foreground mask is got by comparison between object point cloud and background model. Each connected region in foreground mask is defined as a local mask. Ambiguous regions extraction is performed on local mask to extract ambiguous regions to separate adhesive and occluded objects. At last, multiple adjacent superpixel sets are selected to relabel ambiguous regions on pixel-level. In followed parts, we introduce these three modules in more details.

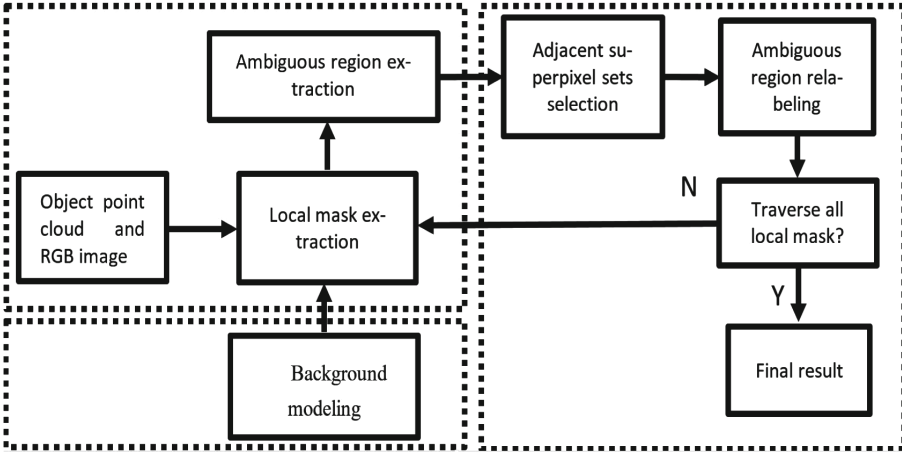


Fig. 2. System overview: background modeling, Ambiguous Regions Extraction, Ambiguous Regions Relabeling.

3.2 Background Modeling

To separate objects from the dusty conveyor belt, a depth-based background subtraction algorithm is used. Most existing methods for background subtraction build a color background model by a number of background frames [14]. However, in industrial scene, camera is fixed while the conveyor belt is running continuously. Color-based model is unreliable in industrial scene as background colors are changing extremely. Instead, depths gained by 3D sensors are more stable, so it is easy to think of using depth information as a cue to achieve background subtraction. Though depth information is more credible than color, a simple subtraction between background point cloud and object point cloud is ineffective. The reasons come that conveyor belt is deform because of the weight of objects and the degree of deformation is not coincident on different parts. Moreover, depth is also influenced by the vibration caused by the running of conveyor belt. So we build a depth-based background Gaussian Mixture Model to remove the background from the image.

Adaptive Gaussian Mixture Model proposed in [7] re-investigates the update equations and utilizes different equations at different phases. This allows algorithm learn faster and more accurately as well as adapt effectively to changing environments. We use depth information to build Adaptive Gaussian Mixture Model of background conveyor belt. Each pixel is modeled by a mixture of K (K is a small number from 3 to 5) Gaussian distributions. We find that $K = 3$ is enough in our scenario by experiment as the depth varies within a certain range. Background conveyor belt frames are inputted to build background model, and when a new object RGB-D image comes, a comparison is processed between the point cloud and the model to generate foreground mask of solid waste objects.

3.3 Ambiguous Region Extraction

In foreground mask, there are several connected regions. Each connected region is defined as a local mask, and it is difficult to judge whether a local mask is single object or touched objects caused by adhesion and occlusion. Furthermore, touched objects separating is also a difficulty. Our algorithm solves these two problems by extract ambiguous regions based on inner edges and SLIC superpixels.

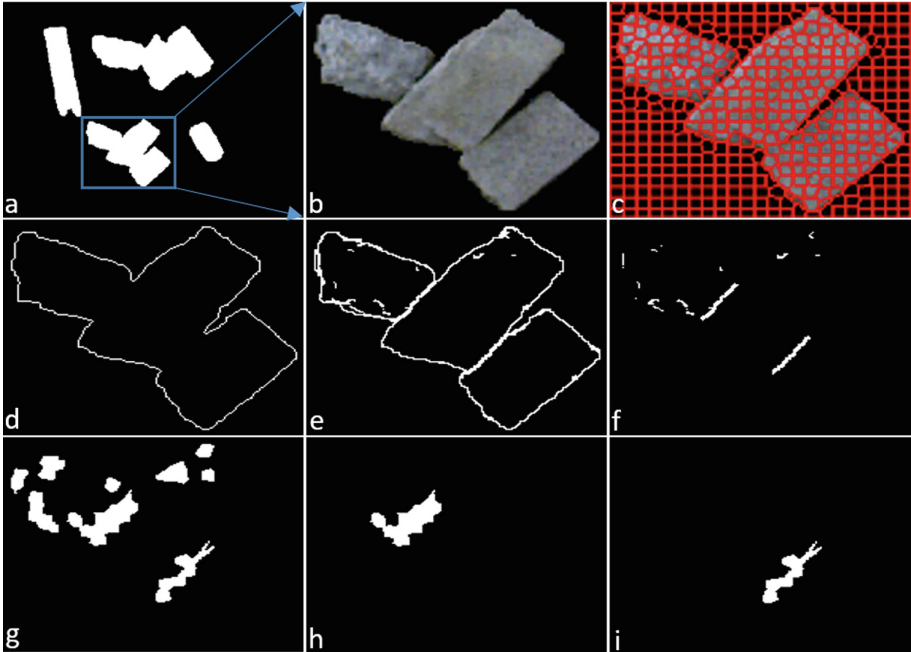


Fig. 3. The process of ambiguous region extraction. (a) Foreground mask. (b) Local image without background. (c) The result of SLIC. (d) The contour of local mask. (e) Local edge map. (f) Local inner edge map. (g) Borderline superpixels. (h) and (i) are ambiguous regions.

Firstly, for each local image, a SLIC superpixel segmentation is performed on and the superpixel set is extracted and defined as $S = \{s_1, s_2, s_3, \dots, s_{n-1}, s_n\}$ based on the result of SLIC. s_i denotes a superpixel and it is a pixel set as it consists of multiple pixels. Local edge map is also generated based on computing depth gradient and defined as E_m . Let F_c denotes the contour of local mask, and then, inner edge map is computed by

$$E_{inner} = E_m - F_c \oplus C_{2*k+1}, \tag{1}$$

where E_{inner} is inner edge map and $F_c \oplus C_{2*k+1}$ denotes a dilation operation on F_c with kernel size of $2 * k + 1$. Inner edge pixel set E_p is extracted based on E_{inner} by

$$E_p = \{p(x, y) | E_{inner}(x, y) = 255\}, \tag{2}$$

where $E_{inner}(x, y)$ means the pixel value of column x and row y in E_{inner} . Then, we combine inner edge pixel set and SLIC superpixel set to extract borderline superpixels by

$$B_{sp} = \left\{ s_k \left| \begin{array}{l} p \in s_k \\ \text{and } p \in E_p \end{array} \right. \right\}, \tag{3}$$

where B_{sp} is a set of borderline superpixels and p is the pixel in image. As shown in Fig. 3 (g), there are several connected borderline superpixels in B_{sp} . Each connected borderline superpixel set is extracted and defined as a borderline region B_{region} . After that, an iteration is used to extract ambiguous region:

$$M_{obj} = M_{local} - B_{region}^{x_{th}}, \tag{4}$$

where M_{local} is local mask and $B_{region}^{x_{th}}$ is borderline region with x_{th} times expansion. $B_{region}^{x_{th}}$ is expanded to $B_{region}^{(x+1)_{th}}$ by

$$B_{region}^{(x+1)_{th}} = B_{region}^{x_{th}} \cup A_{sp}^{x_{th}}, \tag{5}$$

where $A_{sp}^{x_{th}}$ is the set of neighboring superpixels of $B_{region}^{x_{th}}$ and $B_{region}^{x_{th}}$ is borderline region with x_{th} times expansion. The iteration is processing with increasing x from 0 to 4 and stopped when $x > 4$ or M_{obj} has two or more effective parts (have 7 superpixels or more). When the iteration is over, the importance score of borderline region is computed by

$$P_{B_{region}^{y_{th}}} = \begin{cases} \left(1 - \frac{\varphi(B_{region}^{y_{th}})}{\varphi(M_{local})}\right)y & \text{if } f = 1 \\ 0 & \text{if } f = 0 \end{cases}, \tag{6}$$

where $\varphi(B_{region}^{y_{th}})$ denotes the number of pixels in $B_{region}^{y_{th}}$, $B_{region}^{y_{th}}$ is the final borderline region and y is the final times of expansion. f denotes whether extracted borderline region can separate M_{obj} apart or not. $f = 1$ means M_{obj} has two or more effective parts and $f = 0$ mean it do not. If the score is larger than threshold C ($C = 0.4$ in our algorithm), we take this borderline region as an ambiguous region. If a local mask has ambiguous regions, it consists of two or more objects. Otherwise, the local mask is a single object.

3.4 Ambiguous Region Relabeling

A local mask contains touched objects is needed to more precisely segment by assigning every pixel a label. Extracted ambiguous regions separate local mask into multiple object parts, and effective ones are selected as object bodies. As shown in Fig. 4 (a), Pixels in different object bodies are assigned with different labels la ($la = 1, 2, 3 \dots$) and the ambiguous regions and unlabeled parts in OBJ_m are labeled with 0.

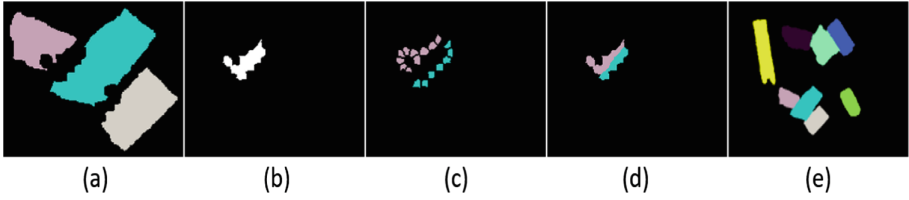


Fig. 4. The process of ambiguous regions relabeling. (a) Labeled object bodies with different labels. (b) An ambiguous region. (c) Selected adjacent superpixel sets. (d) Labeled ambiguous region. (e) Segmentation result of our method. (Color figure online)

For relabeling unclassified pixels ($la = 0$), adjacent superpixel sets are selected as representatives of objects. To an object in construction waste, different parts are not always coincident in feature space. It is not efficient to find a model to represent an object. So the superpixels adjacent to the ambiguous region are selected, and these superpixels have two or more labels. The superpixels with same label are extracted and defined as a superpixel set to represent an object. Two or more superpixel sets are extracted adjacent to an ambiguous borderline region, and then our algorithm classifies the unclassified regions on pixel-level by computing the similarity between a pixel and these sets. For every superpixel, the mean values in LAB color space, depth space and xy space are computed to represent it. So the similarity between a pixel and a superpixel is defined as:

$$d_i = \frac{w_{lab} * d_{lab} + w_{depth} * d_{depth} + w_{xy} * d_{xy}}{w_{lab} + w_{depth} + w_{xy}}, \quad (7)$$

where d_{lab} is the similarity in LAB space, d_{depth} is the similarity of depth space and d_{xy} is the similarity in xy space. These similarities are all computed based on Euclidean distance in each feature space. w_{lab} , w_{depth} and w_{xy} ($w_{lab} = 4$, $w_{depth} = 3$, $w_{xy} = 3$ are chosen by experiment in our algorithm) are the weight of each similarity and i is the serial number of superpixels. Then, the similarity between a superpixel set and a pixel is defined as

$$d = \min_{0 < i <= n} (d_i), \quad (8)$$

where n is the number of superpixels, d is the similarity between a superpixel set and a pixel. Each pixel in ambiguous region is reassigned with the label of most similar (the smaller the value of d , the more similar) superpixel set. After relabeling ambiguous regions, pixels in local mask are all labeled. However, there may be some areas isolated. So we refine the result by finding these independent pixels and reassigning them with the label of most neighbor pixels. After traversing all local masks, all pixels in image are labeled and solid waste objects are all separated as shown in Fig. 4 (e). For seeing single object clearly, different objects are randomly colored.

4 Experiments and Results

To evaluate the performance of algorithm, a representative dataset is needed. However, there is no available construction waste dataset. So we built a scene to simulate the working environment of the robot as shown in Fig. 5 : 3D sensor camera was fixed on a shelf at 1.2m from the conveyor belt and belt was covered with soil and gravel. The objects collected include stone, brick and wood. These objects have irregular shapes and deposition of dusty particles on the surface of them lead to color degradation. We used ASUS Xtion PRO to capture 618 RGB-D images of construction waste. And 100 RGB-D images of dusty conveyor belt were also captured to build Adaptive Gaussian Mixture Model of background. Construction waste dataset contains two kinds of scenes. A simple scene that objects are randomly placed on the belt and every object is isolated, as is shown in Fig. 6 row one. A complex scene that objects are also randomly placed on the belt and some object are adhesive or occluded, as is shown in Fig. 6 row two to five. To evaluate our algorithm, manually labeled ground truths of objects are also provided.

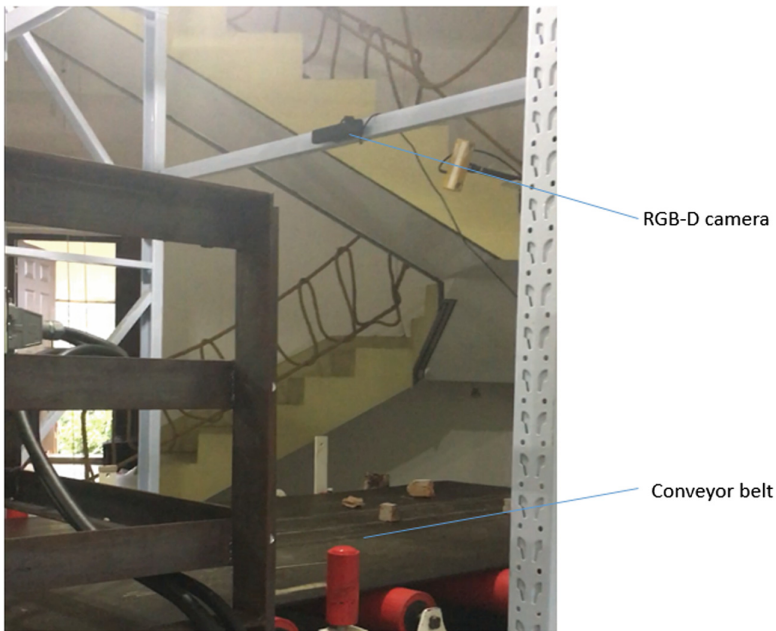


Fig. 5. The simulated scenario of robot working environment.

The performance of our segmentation method is evaluated in terms of how successful it is to segment the objects in the scene. For the task of solid waste sorting, precise object masks which contain positions and contours are important. We define R_i as the pixel set of the i^{th} object in segmentation result,

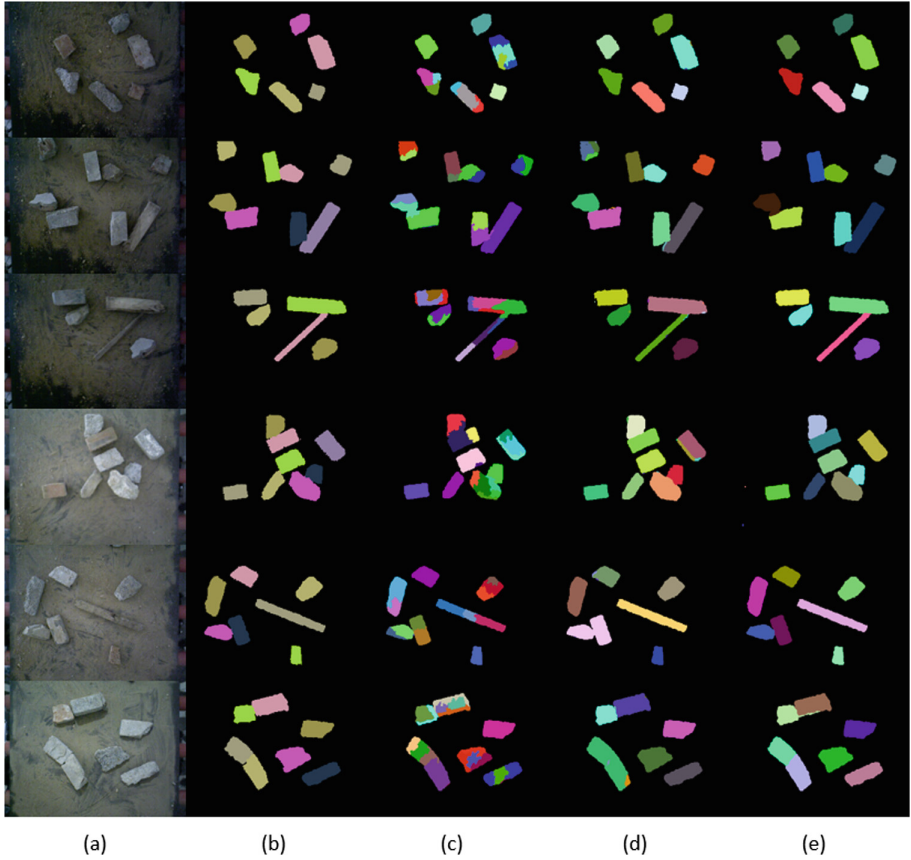


Fig. 6. Some comparison results of different methods. (a) Original images. (b) The ground truths. (c) The results of [15]. (d) The results of [12]. (e) The results of our algorithm. The input for each algorithm are RGB images and point clouds without background. The objects in results are colored randomly. (Color figure online)

G_i is the pixel set of i^{th} object in ground truths and i is the serial number of the objects. So, we analyze the results quantitatively by

$$P_i = \frac{\varphi(I_i) - \varphi(O_i)}{\varphi(G_i)}, \quad (9)$$

where $I_i = G_i \cap R_i$, is the intersection set of G_i and R_i . $O_i = R_i - G_i$, is the pixel set when pixels are in R_i while out of G_i . $\varphi(I_i)$ means the number of pixels in I_i and P_i denotes the segmentation precision of i^{th} object. Over-segmentation and under-segmentation both lead to a low segmentation precision (Fig. 7).

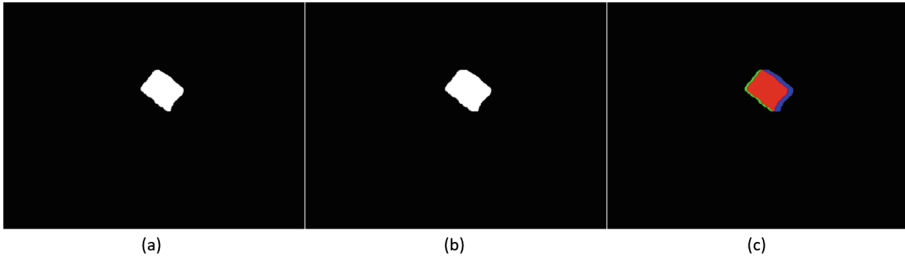


Fig. 7. (a) pixel set of G_i . (b) Pixel set of R_i . (c) Red pixel set is I_i and blue pixel set is O_i . (Color figure online)

We compare our method with other two algorithms of [12] and [15]. The average segmentation precisions of all objects are shown in Table 1. Form Table 1, our method performs better than other two algorithms over the entire dataset. Our segmentation results reach 99.14% segmentation precision in simple scenes and 90.69% segmentation precision in complex scenes. While most objects have been segmented successfully, segmentation error happens (shown in Fig. 5 last row) when some pixels in ambiguous region of two objects are not labeled correctly.

Table 1. Average precision of object segmentation. Compare with [12] and [15].

	Number of images	Number of objects	[12]	[15]	Ours
Simple scenes	345	738	93.73%	32.72%	99.14%
Complex scenes	273	1911	81.82%	41.00%	90.69%
Simple scenes without background	345	738	98.11%	72.83%	99.14%
Complex scenes without background	273	1911	87.98%	64.92%	90.69%

Because of color degradation, color-based algorithms lose their effect. Though algorithm proposed in [15] is newer than [12], it does not achieve a good performance on construction waste dataset as it depends more on color cues. Method presented in [12] computes surface patches by using a mixture of planes and NURBS and performs graph cut on patches to arrive at object hypotheses segmented, so it is plane-based and not affected by interferential color information. But it does not perform well when adhesion and occlusion happen. Our algorithm gets better segmentation results than other two algorithms as color does not play a decisive role in our algorithm and meantime a strategy is proposed to extract ambiguous regions to separate adhesive and occluded objects specially. Masks of segmented objects are provided by our algorithm as they contain necessary information which construction waste sorting needs.

5 Conclusion

In this paper, we have presented a RGB-D based segmentation method for solid waste objects segmentation in cluttered conveyor belt scene. The segmentation results provide the location and boundaries of each solid waste object for the sorting process accurately. In contrast to existing approaches, our method yield satisfactory results in color degraded environment, furthermore, it also performs quite well in the cases where adhesion and occlusion occurred between solid waste objects. Our algorithm deals with adhesion and occlusion by extracting and relabeling on the so-called ambiguous regions to generate accurate segmentation results. To evaluate the proposed algorithm, we have additionally built a dataset of construction waste. The presented results of our method show that it is promising and is looking forward to being used in robotic task of construction waste sorting.

Acknowledgments. This work was supported by Zhejiang Provincial Natural Science Foundation of China under Grant numbers LY15F020031 and LQ16F030007, National Natural Science Foundation of China (NSFC) under Grant numbers 11302195 and 61401397.

References

1. Banica, D., Sminchisescu, C.: Second-order constrained parametric proposals and sequential search-based structured prediction for semantic segmentation in RGB-D images. In: *Computer Vision and Pattern Recognition*, pp. 3517–3526 (2015)
2. Chen, D., Mirebeau, J.M., Cohen, L.D.: A new finlser minimal path model with curvature penalization for image segmentation and closed contour detection. In: *CVPR*, pp. 355–363 (2016)
3. Fu, X., Wang, C.Y., Chen, C., Wang, C., Kuo, C.C.J.: Robust image segmentation using contour-guided color palettes. In: *IEEE International Conference on Computer Vision*, pp. 1618–1625 (2016)
4. Gupta, S., Arbelaez, P., Malik, J.: Perceptual organization and recognition of indoor scenes from RGB-D images. In: *Computer Vision and Pattern Recognition*, pp. 564–571 (2013)
5. Höft, N., Schulz, H., Behnke, S.: Fast semantic segmentation of RGB-D scenes with GPU-accelerated deep neural networks. In: Lutz, C., Thielscher, M. (eds.) *KI 2014. LNCS (LNAI)*, vol. 8736, pp. 80–85. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11206-0_9
6. Holz, D., Holzer, S., Rusu, R.B., Behnke, S.: Real-time plane segmentation using RGB-D cameras. In: Röfer, T., Mayer, N.M., Savage, J., Saranl, U. (eds.) *RoboCup 2011. LNCS (LNAI)*, vol. 7416, pp. 306–317. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32060-6_26
7. Kaewtrakulpong, P., Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection. In: Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S. (eds.) *Video-Based Surveillance Systems*, pp. 135–144. Springer, Heidelberg (2002). https://doi.org/10.1007/978-1-4615-0913-4_11

8. Leonardis, A., Gupta, A., Bajcsy, R.: Segmentation of range images as the search for geometric parametric models. *Int. J. Comput. Vis.* **14**(3), 253–277 (1995)
9. Mishra, A., Aloimonos, Y., Fah, C.L.: Active segmentation with fixation. In: *IEEE International Conference on Computer Vision*, pp. 468–475 (2009)
10. Mishra, A., Shrivastava, A., Aloimonos, Y.: Segmenting simple objects using RGB-D. In: *IEEE International Conference on Robotics and Automation*, pp. 4406–4413 (2012)
11. Rao, D., Le, Q.V., Phoka, T., Quigley, M.: Grasping novel objects with depth segmentation. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2578–2585 (2010)
12. Richtsfeld, A., Mrwald, T., Prankl, J., Zillich, M., Vincze, M.: Segmentation of unknown objects in indoor environments. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4791–4796 (2012)
13. Silberman, N., Sontag, D., Fergus, R.: Instance segmentation of indoor scenes using a coverage loss. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8689, pp. 616–631. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_40
14. Sobral, A., Vacavant, A.: A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Comput. Vis. Image Underst.* **122**, 4–21 (2014)
15. Toscana, G., Rosa, S.: Fast graph-based object segmentation for RGB-D images. *arXiv preprint arXiv:1605.03746* (2016)
16. Wang, Z., Liu, H., Wang, X., Qian, Y.: Segment and Label Indoor Scene Based on RGB-D for the Visually Impaired. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) *MMM 2014*. LNCS, vol. 8325, pp. 449–460. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04114-8_38