3D Complex Scenes Segmentation from a Single Range Image Using Virtual Exploration

P. Merchán[1], A. Adán[2], S. Salamanca[1], C. Cerrada[3]

[1]Escuela de Ingenierías Industriales, Universidad de Extremadura, Avda. de Elvas s/n, 06071 Badajoz, Spain
{pmerchan,ssalaman}@unex.es
[2]Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ronda de Calatrava 5, 13071 Ciudad Real, Spain
aadan@inf-cr.uclm.es
[3]ETSI Industriales, UNED, Ciudad Universitaria s/n, 28040 Madrid, Spain
ccerrada@ieec.uned.es

In this paper we present a method for automatic segmentation of 3D complex scenes from a single range image. A complex scene includes several objects with: irregular shapes, occlusion, the same colour or intensity level and placed in any pose. Unlike most existing methods which proceed with a set of images obtained from different viewpoints, in this work a single view is used and a 3D segmentation process is developed to separate the constituent parts of a complex scene. The method is based on establishing suitable virtual-viewpoints in order to carry out a new range data segmentation technique. For a virtual-viewpoint a strategy [3D range data] - [2D projected range data] - [2D segmentation] - [3D segmented range data], is accomplished. The proposed method has been applied to a set of complex scenes and it can be said that the results guarantee the benefits of the method.

Introduction

Image segmentation is one of the most important subjects in image processing which finds wide applications in pattern recognition and 3D vision. It consists of partitioning the image into its constituent parts and extracting these parts of interest (objects). Until now a wide variety of different segmentation algorithms have been developed. Criteria used in the image-partitioning process are largely dependent on the nature of the input data, the desired high-level task and the nature of the scene.
In complex scenes, there are usually occluded surfaces in the image obtained from a view point, that is why most of the techniques developed until now use sets of data (intensity images or range images) taken from different viewpoints to obtain a complete model of the scene [1], [2], [3], [4], [5]. In these techniques, the problem of the camera positioning to reduce the number of views, known as the *Best-Next-View* problem, arises.

When we say "3D segmentation", two environments can be referred to single scenes or complex scenes. In the first case, 3D segmentation involves isolated objects, scenes with no occlusion, intensity image segmentation techniques of stereo pairs, etc.

Segmentation means to extract features or primitives of the object. Different strategies are applied to perform these tasks. In general, approaches can be categorized into two types: edge-based approaches and region-based approaches. In edge-based approaches, the points located on the edges are first identified, followed by edge linking and contour processes. Edges or contours could segment the scene. In region-based approaches a number of seed regions are first chosen. These seed regions grow by adding neighbour points based on some compatibility threshold [6], [7], [8], [9]. In [10] a segmentation approach of range images is proposed. They use curved segments as segmentation primitives instead of individual pixels. So the amount of data is reduced and a fast segmentation process is obtained. A simulated electrical charge distribution is used in [11] in order to establish the surface curvature of the objects.

3D Segmentation of complex scenes is a bigger problem in computer vision. In the worst case, a complex scene includes: objects with irregular shapes; objects viewed from any direction, objects with self-occlusion or partially occluded by other objects and uniform intensity/colour appearance. In our case, we are concerned with complex scenes and range images to solve this problem. Range images have been used most frequently in 3D object recognition tasks and a lot of progress has been made in this field. Although several techniques based on modelling have been applied to segment parts of the scene using range data [12], [13], [14], [15], there are very few researchers working on segmentation/recognition based on range data processing. Therefore nowadays it is widely accepted that recognition of a real world scene based on a single range view is a difficult task.

In this paper, we present a region-based segmentation method for partitioning a complex scene image into meaningful regions (objects). To do this, we use a single range data image. In the next section, we will have a glance at the whole process. In section 3 and section 4, we describe two main stages: scene exploration and scene segmentation, respectively. Next, in section 5, experimental results achieved with the application of the proposed method to a set of real range images are shown. Finally, in section 6 we conclude and discuss limitations and future research.

2 Overview of the Process

As it has been said, we use a single view of the scene for separating all its constituent objects. The proposed segmentation scheme is an iterative process composed of three successive steps:

1. *Range data obtaining*: We use real range images obtained with a gray range finder. For the first iteration, range data are the scene surface points given by the sensor. For following iterations, the new range data will be the old range data without the range data segmented in the last iteration.
2. *Scene exploration*: A virtual camera is placed in the scene for exploring the range data and a procedure for searching for an appropriate virtual viewpoint is accomplished. The strategy developed to choose these viewpoints will be explained in section 3.

3. *Scene Segmentation*: taking into account a virtual viewpoint, 2D data orthogonal projection is taken to perform a segmentation process. When the segmentation of the processed 2D image has been finished, an inverse transformation is accomplished to reassign each 2D segment to its corresponding 3D segment in the scene.

The process is iteratively executed until there are no objects in the scene. For each iteration several possibilities can be given:

- No segmentation. It means that for the current exploration viewpoint it is not possible to segment any part of the current scene data.
- Segmentation. The current scene is segmented into several parts. For each segmented part there are two possibilities :
  - The segmented part corresponds to an object.
  - The segmented part corresponds to more than one object. In this case each segment will be considered as a new range image for step 1.

3 Scene Exploration

In this section we will explain how the election of the virtual viewpoint is made. We use the mesh of nodes created by tessellating the unit sphere in order to limit the number of viewpoint candidates. Nowadays we have considered a tessellated sphere formed by 320 nodes with 3-connectivity (see Fig. 1.a) where each node N defines the viewpoint ON, O being the centre of the sphere. As it has been said before, we are interested in the projected image of the range data over the viewpoint chosen. Since a viewpoint and its opposite provide the same projected image of the scene we only consider the half-tessellated-sphere.

Probability between 0 and 1 is associated to each node N according to the topological organization proposed in [16], [17]: the *Modeling Wave Set*, MWS. Before explaining the probability mapping procedure, a short reference about MWS will be given.

MWS structure organizes the nodes of the tessellated sphere as disjointed subsets, each one containing a group of nodes spatially disposed over the sphere as a closed quasi-circle (see Fig. 1.b). Each of the disjointed subsets is called *wave front*, WF. To build a *modelling wave* MW, a node of the tessellated sphere must be chosen as the origin of the structure. This node, called *initial focus*, constitutes the first *wave front* WF_1 of the MW and will be used to identify it. Then, the remaining WF are sequentially obtained by building rings of nodes over the sphere until it is completely covered.

MW structure is used for updating the probabilities P associates to the nodes (viewpoints) for each iteration. For the t iteration, we select the node with the highest value of probability as the appropriate viewpoint,. The viewpoint chosen defines an initial focus $WF_1(t)$ and its corresponding MW(t) (Fig. 1.b). After the process and taking into account the result of the segmentation, the map of probabilities is updated for iteration t+1as follows:

– $P(WF_1(t+1))=0$. The old focus probability is assigned a value 0 because we will not use this node (or viewpoint) any more. It does not matter if we have segmented regions or not.

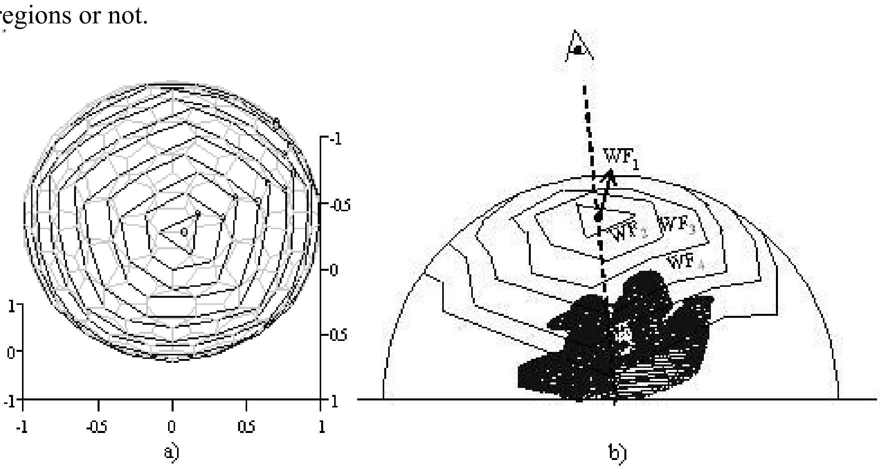


Fig. 1. a) Tessellated sphere and a MW drawn over it. b) Exploration viewpoint defined by a node of the sphere and the associate MW structure

– If the iteration is successful (segmented regions), we consider the nodes around the focus as *good* points to be selected for future segmentations and we increase the probability associated to the closest neighbours of WF_1. The nearest WF to WF_1 will have the highest increase. In this case the expression used for updating is the following:

$$P(WF_i(t+1)) = \frac{P(WF_i(t)) + (1-0.2i)}{v_{\max}}, \quad i = 2,3,4,5 . \tag{1}$$

where v_{max} is the maximum value among the obtained after summing the term (*1-0.2i*) to the old probabilities.

– If the iteration is not successful (no segmented regions), we consider the nodes around the focus as *bad* points to be selected for future segmentations. So we decrease the probability associated to the close neighbours of WF_1. The nearest WF to WF_1 will have the highest reduction. The expression used in this case is:

$$P(WF_i(t+1)) = \frac{P(WF_i(t)) - (1-0.2i)}{v_{\max}}, \quad i = 2,3,4,5 . \tag{2}$$

When the probability values have been changed, the t+1 iteration begins and the node with the highest value of probability is again selected as the appropriate viewpoint. Figure 2 illustrates the general procedure. Then the process continues with the next step, the scene segmentation.

For the first iteration, the map of probabilities must be imposed in an arbitrary manner because scarce information about *good* or *bad* viewpoints is known beforehand. It is just known that the viewpoint (or node) defined by the camera N_c is a *bad* viewpoint and consequently its neighbour nodes should be as well. Therefore, we can model this situation as the 0 iteration where: all nodes have probability 1, the viewpoint chosen is N_c and it is not successful. So in the first iteration $P(N_c)=0$ and N_c neighbour probabilities will be close to zero. Next an arbitrary node with probability 1 will be selected as the first viewpoint and the whole process will be run.

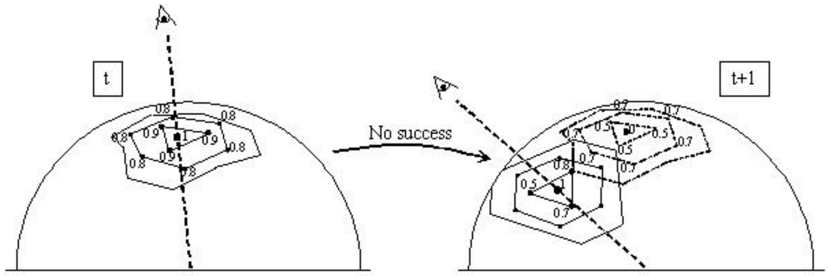


Fig. 2. Probability distribution and updating after a iteration step

4 Scene Segmentation

To obtain the constituent objects of the scene, we perform a transformation of the 3D data and the segmentation is made on a 2D image. The sequence developed to reach this objective is explained in Fig. 3.a).

The process begins with a 3D →2D transformation using the viewpoint chosen in the previous step. This conversion provides a multidimensional structure that stores the information of the relationship between every 3D point and its respective 2D pixel. We call this structure *multipixel matrix* (see Fig. 3.b). It is an array of two-dimensional matrixes. The first matrix stores the indexes of the projected points corresponding to the 2D black pixels. Since there may be points whose projection match up to the same 2D pixel, we must accumulate the information of each one in order not to lose any of them when the inverse transformation is made. To do this, we develop the multipixel matrix in such a way that each vector in the third dimension (*depth* in Fig. 3.b) stores all the 3D points with identical 2D projection.

With the projected points I_p we conform a 2D image with the purpose of carrying out a specific image processing. So 2D image size and a conversion to a binary image I_1 are given. Then we run a 2D region-based segmentation algorithm over the binary image. If the viewpoint is appropriate several binary disjointed segments of the image will be obtained and the algorithm will continue. On the contrary, if there are no disjointed segments, the result will be negative and a new viewpoint must be selected. We call I_2 the 2D segmented image.

After the segmentation phase, we define the regions on the original projected image from these segments I_{sp}. The binary conversion is undone and the regions on the recovered 2D image are found. Once we have segmented regions on the 2D projection, the information stored in the multipixel matrix is used to carry out the inverse transformation. This way, if the algorithm has segmented disjointed regions in the projected image, the corresponding 3D points will be extracted at the end of the iteration. Each disjointed region is considered as the viewed part of an object in the scene and consequently as a 3D segment.

The segmented parts are removed from the scene for the next iteration and a new scene is again explored. The process continues searching a new viewpoint with the remaining data. At the same time and in the same way, each segmented part could become a new range image for exploring if the number of data-points of the segment is high enough. So a segmentation distributed strategy could be achieved if necessary.

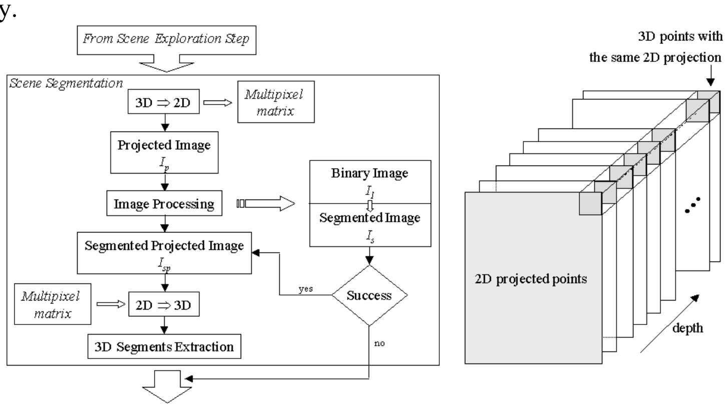


Fig. 3. a) Scene segmentation chart. b) Multipixel matrix structure

Fig. 4 shows an example illustrating the successive phases in the segmentation process. It begins at the first iteration. In a) intensity image and the corresponding range image of the scene are shown. An exploration viewpoint is chosen following section 3 and the corresponding projected range-data are plotted in b). After 2D image processing we deal with the image I_1 and perform the segmentation. As it can be seen in c) several segments are extracted. Then we recover the 3D points corresponding to each segment (d) and we identify them. Finally these points are removed to the original range image and the new range data for the next iteration is shown in e).

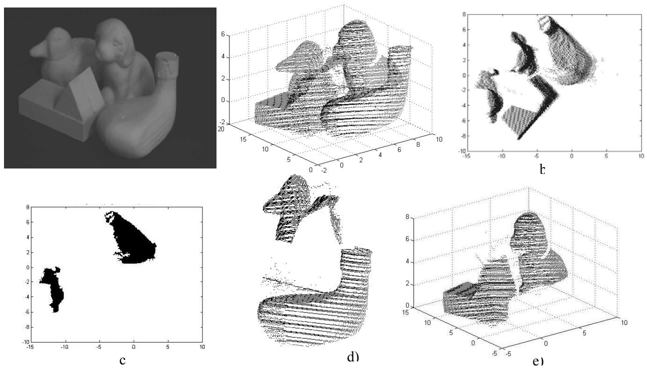


Fig. 4. Segmentation process. Example

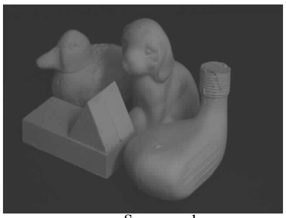
5 Experimental Results

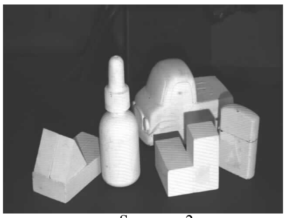
We have tested our algorithm on a set of real data images. The scenes are composed of several objects that can have irregular shapes. The objects are viewed from any direction and there are objects with self-occlusion or partially occluded by other objects. Moreover, the objects have been painted with the same colour.

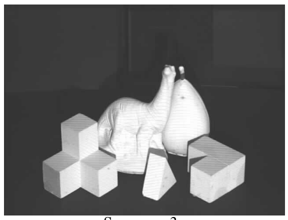
The results achieved with some of the scenes have been summarized in Table 1. The intensity images of the scenes numbered in the first column as *Scene no.1* up to *Scene no. 5* can be seen in Fig. 5. Table 1 illustrates, in the second column, the total number of points that compose the input range images, which has been called N_T. In the third column the number of objects that constitute the scenes (*Ob*) are shown. The number of iterations of our algorithm needed to separate all the objects in the scenes, denoted *I*, appears in column four. In the next column, we have registered the number of points belonging to each segmented 3D object once the process has terminated, N_{Ob}. The last column reveals the quantity of points that has been lost during the process showing the percentage with respect to the initial data. As it can be seen, this number is low enough to confirm the goodness of our method.

Table 1. Results presentation

	N_T	*Ob*	*I*	N_{Ob}					N_L *(%)*
				Ob1	*Ob2*	*Ob3*	*Ob4*	*Ob5*	
Scene no. 1	29932	4	4	10520	4234	8439	6527	-	0.71
Scene no. 2	24632	5	3	5014	6177	3349	5047	4905	0.57
Scene no. 3	19661	5	3	5453	3131	3638	5821	1424	0.98
Scene no. 4	16737	5	2	3327	3027	4008	3294	2786	1.5
Scene no. 5	22936	4	18	6539	3746	8778	2822	-	4.58


Scene no. 1


Scene no. 2


Scene no. 3


Scene no. 4

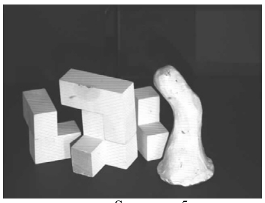

Scene no. 5

Fig. 5. The five scenes analyzed in table 1

Figs. 6 and 7 exemplify the results achieved after the segmentation algorithm has been run on two scenes. In Fig. 6.a) we show the intensity image of scene number 2.

The range image obtained with the range finder of such a scene is shown in b). This range image constitutes the input data to the segmentation process. Fig. 6.c) displays the viewpoint selected among the possible candidates and in (d) the projected image obtained with this viewpoint can be seen. Fig. 6.e) shows three objects segmented at this iteration. In Fig. 6.f) we show the new scene to be analyzed after the elimination of the segmented objects and the viewpoint that gives the projected image shown in Fig. 6.g). Range data of each segment recovered at the end of the iteration is exposed in Fig. 6. h).

As it mentioned in section 4, sometimes the extracted segment corresponds to more than one object of the scene. This situation is contemplated by the method in the following manner: the algorithm automatically detects that such circumstances have taken place and performs a new segmentation over those segments starting from *Scene Exploration*. The probability values considered to select the viewpoint are those existing in the iteration in which the region was segmented. This occurred, i.e., in scene number 3. The four objects segmented after a number of iterations are shown in Fig. 7. b) to e). As it can be seen the segment in e) corresponds to two objects. Fig. 7.f) and g) illustrate the two objects obtained when the algorithm continues its execution recursively.

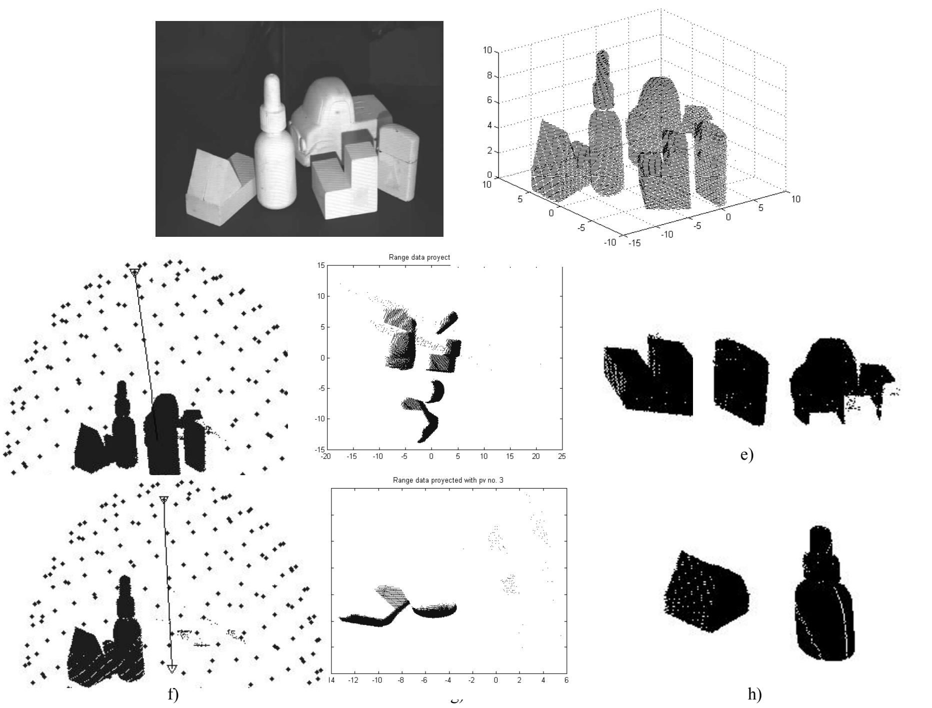


Fig. 6. Input data image and resultant segmentation. a) Intensity image of scene no.2. b) Range image of the same scene. c) Selected viewpoint among the possible candidates. d) Projected image obtained. e) Segmented object no.1, 2 and 3. f) Resultant scene after the segmentation and new viewpoint. g) Corresponding projected image. h) Segmented object no.4 and 5.

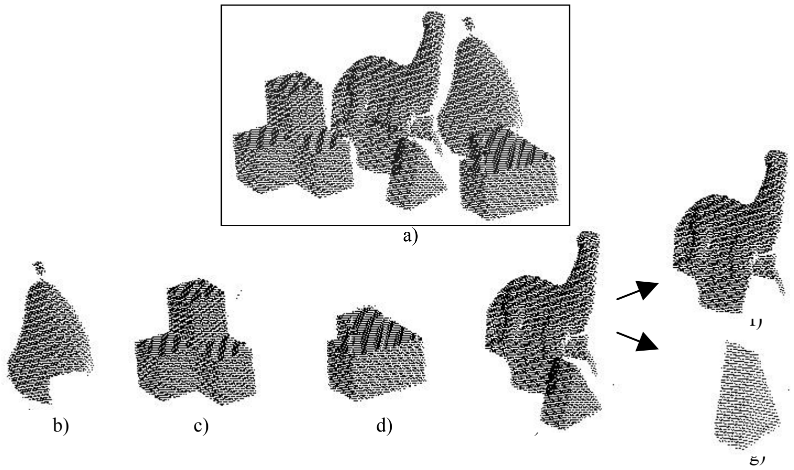


Fig. 7. Input data image and resultant segmentation. a) Range image of scene no. 3. b) to d) Segmented objects no. 1 to 3. e) Segmented region composed of two objects. f) and g) Segmented regions no. 4 and 5

6 Conclusions

In this paper, a method for automatic segmentation of 3D complex scenes has been presented. Contrasting most existing techniques, which proceed with a set of images obtained from different viewpoints, an important feature of our approach is that we use a single range image to separate the constituent parts of a 3D complex scene, which can include several objects with irregular shapes, occlusion, the same colour and place in any pose. This is achieved by applying a new strategy based on the selection of virtual-viewpoints that let us develop a range data segmentation algorithm over 2D projected images. Next, an inverse transformation is executed to relocate each 2D segment to its corresponding 3D object in the scene.

Experiments carried out on a set of real range images have proved the validity of our method. They have shown that it can successfully be used to perform the segmentation of a 3D scene. Nowadays we are improving this strategy for scenes with a higher number of objects and occlusion complexity.

References

1. J.E. Banta, Y. Zhien, X. Z. Wang, G. Zhang, M.T. Smith, and M.A. Abidi, "A "Best-Next-View" Algorithm for Three-Dimensional Scene Reconstruction Using Range Images", *Proceedings SPIE,* vol. 2588, pp. 418-429, 1995.

2. R. Pito, “A Solution to the Next Best View Problem for Automated CAD Acquisition of Free-form Objects Using Range Cameras”, *Proceedings SPIE Symposium on Intelligent Systems and Advanced Manufacturing*, Phila, PA, October 1995.
3. Brian Curless and Marc Levoy, ”A Volumetric Method for Building Complex Models from Range Images”, *SIGGRAPH96, Computer Graphics Proceedings*, 1996.
4. Nikolaos A. Massios and Robert B. Fisher, “A Best Next View Selection Algorithm Incorporating a Quality Criterion”, *Proceedings of the 6th British Machine Vision Conference*, pp. 780-789,1998.
5. Michael K. Reed and Peter K. Allen, “Constraint-Based Sensor Planning for Scene Modeling”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no.12, pp. 1460-1467, December 2000.
6. Y. J. Zhang, ”Evaluation and comparison of different segmentation algorithms”, Pattern Recognition Letters, vol. 18, Issue 10, pp. 963-974, 1997.
7. Adam Hoover, Gillian Jean-Baptiste, Xiaoyi Jiang, Patrick J. Flynn, Horst Bunke, Dmitry B. Goldgof, Kevin Bowyer, David W. Eggert, Andrew Fitzgibbon, and Robert B. Fisher, “An Experimental Comparison of Range Images Segmentation Algorithms”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no.7, pp. 673-689, July 1996.
8. Ioannis Stamos and Peter K. Allen, “3-D Model Construction using Range and Image Data”, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, vol. I, pp. 531-536, South Carolina, June 2000.
9. Jianbing Huang and Chia-Hsiang Menq, “Automatic Data Segmentation for Geometric Feature Extraction From Unorganized 3-D Coordinate Points”, *IEEE Transactions on Robotics and Automation,* vol. 17, no. 3, pp. 268-279, June 2001.
10. Xiaoyi Jiang, H. Bunke, U. Meier. “High Level Feature Based Range Image Segmentation”, *Image and Vision Computing* 18 , 817 – 822, 2000.
11. Kenong Wu and Martin D. Levine, “3D part Segmentation Using Simulated Electrical Charge Distributions”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no.11, pp. 1223-1235, November 1997.
12. R. Benlamri, ”Range Image segmentation of scenes with occluded curved objetcs”*, Pattern Recognition Letters* 21 , 1051 – 1060, 2000.
13. Kil-Moo Lee, Peter Mee, and Rae-Hong Park, “Robust Adaptative Segmentation of Range Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no.2, pp. 200-205, February 1998.
14. Johnson A. and Hebert M. “Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes” , *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp 433-449, May 1999.
15. Éric Marchand and François Chaumette, “ Active Vision for Complete Scene Reconstruction and Exploration” , *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, January 1999.
16. A. Adán, C. Cerrada and V. Feliu, “Modeling Wave Set: Definition and Application of a new Topological Organization for 3D Object Modeling”, *Computer Vision and Image Understanding*, vol. 79, pp. 281-307, 2000.
17. A. Adán, C. Cerrada and V. Feliu,”Automatic pose determination of 3D shapes based on modeling wave sets: a new data structure for object modeling”, Image and Vision Computing, vol. 19, pp. 867-890, 2001.