SEGMENTATION OF OBJECT REGIONS USING DEPTH INFORMATION

Nak H. Kim and Jai Song Park

School of Electronics and Information Engineering
Hankuk University of Foreign Studies
Yongin-si Mohyun-myun, Kyunggi-do, Korea
nhkim@hufs.ac.kr

ABSTRACT

This paper describes an object segmentation technique using 3D information extracted by stereo vision. The proposed algorithm is based on the assumption that all objects are lying on a common base plane. From the input stereo images, depth data is first extracted on edge points, and the recovered 3D points are projected onto the base plane. Using the fact that projected points from the same object tend to form a cluster on the base plane, each object region is separated. We demonstrate the efficacy of the algorithm using experimental images.

1. INTRODUCTION

Image segmentation is a process to divide a given image into a set of meaningful regions. There have been a large number of methodologies proposed for this task, each of which has a specific purpose suitable for some application tasks [1]. In this paper, we present a segmentation algorithm that separates each object region in a scene containing unknown number of objects. To be more specific, the algorithm is aimed at dealing with the following working conditions, which typically arise in the context of object recognition:

- Objects are located on a common base plane.
- There is some space between each pair of objects.
- Each object has a relatively simple shape without protruded part extended to other object's area.
- There are no restrictions on colors of objects, and illumination.

Suppose that 3D depth data is recovered for objects in a scene satisfying the above conditions, and that the recovered 3D object points are projected onto the base plane. Then the points from the same object will tend to form a cluster on the base plane. By partitioning projected points into a number of clusters, it will become possible to detect the region occupied by each object. Based on this observation, we have developed a segmentation technique that divides an image into a set of object regions. In order to obtain depth data, an area-based stereo matching algorithm is applied to input stereo images. The 3D coordinates of matched points are computed using measured disparity values, and points are projected onto the base plane. Finally, projected points are clustered into object regions.

The proposed segmentation technique has a number of important application areas. First of all, the segmentation process can be used as a pre-processing step for object recognition. Some recognition algorithms [2, 3] operate under the condition that there is only one object within the scene. For this class of algorithms, the proposed technique can be utilized to segment the input image into a set of sub-images containing separate objects. Secondly, the proposed technique can provide 3D bounding volume parameters for each region occupied by objects, and such information can be utilized for various 3D vision tasks including object recognition, autonomous navigation, and surveillance.

Although there have been a large collection of researches on obtaining 3D depth information from stereo images [4], a relatively few efforts have been addressed to utilize such data for building a higher level scene description. The proposed work is an attempt to build a 3D scene description suitable for higher level processing.

There have been some previous attempts [5, 6] for projecting 3D information onto a certain plane to build a higher level scene description. The novelty of our approach is that we devised an object segmentation algorithm based on the projection method. In particular, our algorithm can find object boundaries even when the number of objects in the scene is not known. In this paper, we present experimental results using a set of indoor stereo images illustrating the performance of the algorithm.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. COMPUTING PROJECTION ONTO A 3D PLANE

2.1. Extraction of 3D data from stereo images

Among various means for obtaining depth information from images, stereo vision is one of the most effective methods. There have been a large amounts of research efforts for improving stereo matching technique [4], and now it is possible to extract a reasonably good dense depth data in practical applications. However, most of currently available algorithms have a limitation in providing reliable depth data in featureless regions, and often fail to produce a usable depth map in indoor environments containing less textured objects, such as table, chair, cup, etc. However, even for such featureless objects, it is still possible to obtain depth data along the boundaries of the objects. In addition, edge or line data is exploited as the input for many high-level vision algorithms.

Since we are interested in dealing with less textured objects in this research, we have chosen an edge-based approach for stereo matching. Thus we will assume that disparity data is available on the edges of one of the input images. Such disparity data can be extracted using an edge-based matching algorithm [1, 5], or more practically, it is also possible to exploit an area-based matching algorithm [4] to get disparities on edge points.

2.2. Projection of 3D points onto the base plane

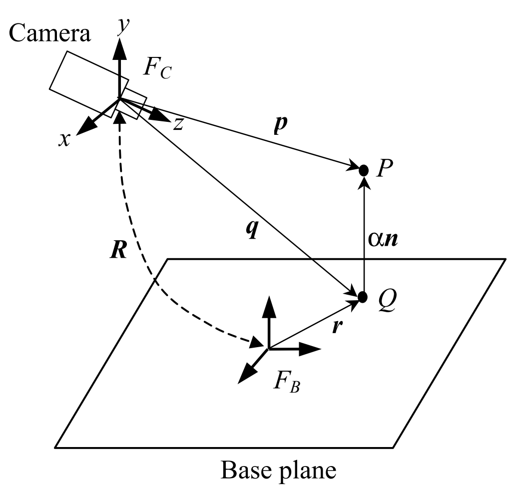


Fig. 1. Projection of a 3D point to a plane.

Let F_C be a camera-centered 3D coordinate frame as shown in Fig. 1, and consider a 3D point *P* represented as $\mathbf{p} = (x \quad y \quad z)^T$ in F_C. Let a base plane Ω be given with a normal vector $\mathbf{n}$. A point $\mathbf{x}$ on Ω satisfies

$$\mathbf{n} \cdot \mathbf{x} = d \tag{1}$$

where d is a plane parameter. The vector equation of the projection Q of P onto Ω is given by

$$\mathbf{q} = \mathbf{p} - \alpha\mathbf{n} \tag{2}$$

where α is a scale factor. If the base plane parameters are known, then for each 3D point $\mathbf{p}$, the scale factor α and $\mathbf{q}$ can be solved simultaneously using the above two equations.

In order to find the projection onto the base plane, the vector $\mathbf{q}$ needs to be converted into a vector $\mathbf{r}$ on the object-centered coordinate frame F_B. This conversion is a transformation between two 3D coordinates, and since the origin of F_B can be set arbitrarily, only the rotation part is sufficient for defining the transformation $\mathbf{r} = R\mathbf{q}$. While there are numerous ways to define the base plane coordinates, a simple and good choice is to define $\mathbf{n}$ as the vertical axis. One such selection leads to the following rotation matrix:

$$R = (\mathbf{u}_z \times \mathbf{n} \quad \mathbf{n} \quad \mathbf{n} \times \mathbf{u}_z \times \mathbf{n})$$

where $\mathbf{u}_z = (0 \quad 0 \quad 1)^T$.

3. FINDING OBJECT AREAS

3.1. Object area detection

The 3D points projected onto the base plane tend to form clusters, each of which corresponds to an object area. In order to exploit this property for object segmentation, a *density image* is defined on the lattice formed by dividing the base plane into equally sized cells. Each element of the density image denotes the number of projected points. The image is constructed by counting the number of matched points projected on the corresponding cell on the base plane.

Figs. 2 and 3 show a pair of stereo images and the corresponding density image. In Fig. 3, the darkness of each point encodes the number of projected points. It can be seen that each object region is formed by a group of roughly aggregated points, and that the object segmentation problem amounts to a clustering procedure on the density image. Under the working conditions of this study, some properties of the density image can be summarized as follows:

1) Feature points from the same object tend to be projected onto nearby locations, and the feature density is high on the object region.
2) Since it has been assumed that some empty space exists between each pair of objects, feature regions will be separated.

3) The number of objects present in the scene is not known a priori.

As far as the clustering algorithm is concerned, although there may be several alternatives to choose from, most clustering techniques are not adequate for our purpose since the number of objects is not known in advance.

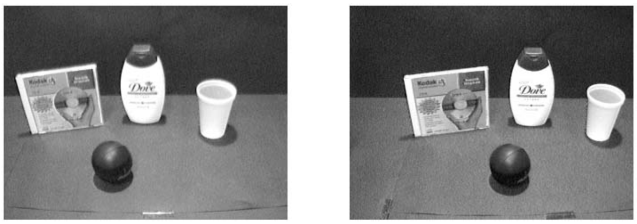


Fig. 2. A stereo image pair

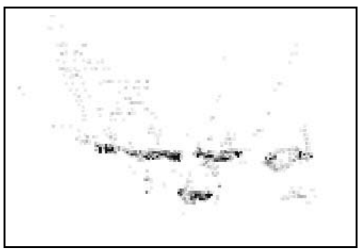


Fig. 3. Projection of matched edge points onto the base plane.

In our experiment, we found that an image-based segmentation algorithm based on simple thresholding and connected component analysis techniques is most effective. First, the density image constructed from projected 3D points is smoothed using local filtering to reduce noise effects. Then a threholding is applied to the density image, and each object region is detected using the connected component analysis technique.

While object regions detected on the density image are areas on the base plane, it is straightforward to extend the region information for segmenting edges on the original image. In order to segment edges, each edge point is projected again onto the base plane, and using the region information, it is possible to segment the edge point into each object region.

3.2. Merging regions

Although the proposed segmentation algorithm described in the previous section has been found to work well for textured objects, a problem has appeared when it is applied to objects with little texture on their surfaces. Consider the plastic cup in the right side of Fig. 2. On the projection image in Fig. 3, it can be seen that regions corresponding to the cup appear as two regions instead of one. This is due to the sparseness of disparity information along horizontal edge segments.

To deal with the problem, we exploited the fact that there exist some horizontal edge segments joining the regions arising from the same physical object. For this purpose, after the initial segmentation is done on the density image, horizontal edge segments are examined. If two regions of similar depth have more than a couple of horizontal edge segments connecting them, two regions are merged.

4. EXPERIMENTAL RESULTS

In this section, we present experimental results applied to real stereo images. The size of each image is 320x240, and the maximum horizontal disparity is 60 pixels. Edges in the image were detected using Laplacian-of-Gaussian edge detector. Stereo matching was performed on each edge point using a correlation-based method [7].

In this experiment, the plane equation of the base plane is assumed known. We observed that the segmentation algorithm is not so sensitive to the small errors in the plane parameters.

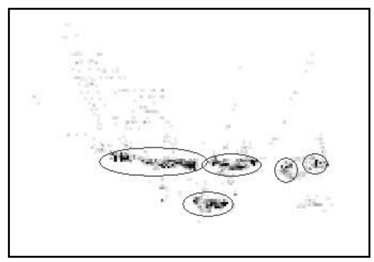
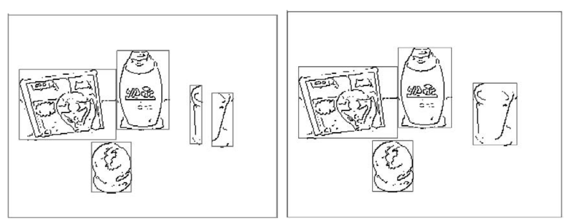


Fig. 4. Detected object regions



(a) Before region merge (b) After region merge

Fig. 5. Segmentation result.

Fig. 4 shows the detected regions marked as ellipses superimposed on the density image for stereo image pair in Fig. 2. Although there are four objects in the scene of Fig. 2, there are five detected object regions in Fig. 4, because the region occupied the plastic cup in the right side is split into two regions due to the reason explained in Section 3.2. In order to correct the error, the region merge process is applied, and in Fig. 5, bounding boxes for object regions are shown overlaid on the edge map. In Fig. 5, only those edge points matched through stereo matching are shown. Since matches could not be obtained on horizontal edges, some edge segments are disconnected. Note that the plastic cup region is detected correctly in Fig. 5(b).

Fig. 6. Another stereo image pair

Another experimental stereo image pair is shown in Fig. 6, where some objects are occluded in monocular image. However, since the objects are located separately in depth, the projections of object points are clearly separated on the base plane as shown in Fig. 7, where detected object regions are marked using ellipses. Although there is an erroneous region in the background, all six objects in the scene are detected correctly. The final segmentation result is shown in Fig. 8.

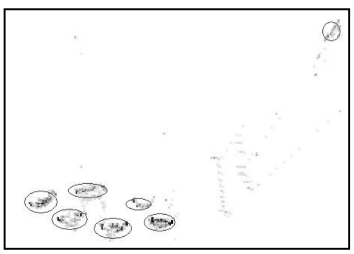


Fig. 7 Detected object region

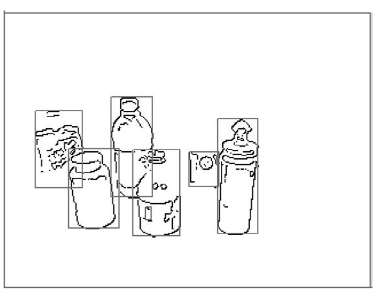


Fig. 8. Segmentation result for Fig. 6.

5. CONCLUSION

In this paper, we have described an object segmentation technique using depth data obtained from stereo images. Based on the property that projections of points from the same object tend to form a cluster on the common base plane, the proposed algorithm finds the object boundary using an image segmentation algorithm. In order to deal with a problem arising in featureless object regions, a region merge process is applied after the initial segmentation is done. The efficacy of the approach is demonstrated using real stereo images.

The proposed segmentation technique has a number of important application areas. First of all, the segmentation process can be used as a pre-processing step for object recognition. Otherwise, the segmentation algorithm can help reduce computation time of an object recognition system by providing initial segmentation information. Secondly, the proposed technique can provide 3D bounding volume parameters for each region occupied by objects, and such information can be utilized for describing the volume occupancy information in various 3D vision applications.

In this research the surface parameters of the base plane is assumed known. We expect that the parameters can be estimated automatically using extracted depth information. However, we have observed that the performance of the segmentation process is not affected by small errors in surface parameters. Thus, for practical purpose, a rough initial guess on the parameters will suffice for most applications.

6. REFERENCES

[1] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2003.

[2] H. Murase and S. K. Nayer, "Visual learning and recognition of 3D objects from appearance," *Int. J. Computer Vision*, 14(1), pp. 5-24, 1995.

[3] S. Sclaroff and A. P. Pentland, "Modal matching for correspondence and recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(6), pp. 545-561, 1995.

[4] D. Sharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Computer Vision*, vol. 47, pp. 7-42, 2002.

[5] O. D. Faugeras, *Three-Dimensional Computer Vision*, MIT Press, 1993.

[6] Y. Sumi et al, "3D object recognition in cluttered environments by segment-based stereo vision," *Int. J. Computer Vision*, 46(1), pp. 5-23, 2002.

[7] H. Hirschmüller, "Improvements in real-time correlation-based stereo vision," In *Proc. IEEE Workshop on Stereo and Multi-Baseline Vision*, 2001.