Journal of Korea Multimedia Society Vol. 11, No. 12, DECEMBER 2008(pp. 1658-1667)

An Approach to 3D Object Localization Based on Monocular Vision

Sunghoon Jung†, Dowon Jang††, Minhwan Kim†††

ABSTRACT

Reconstruction of 3D objects from a single view image is generally an ill-posed problem because of the projection distortion. A monocular vision based 3D object localization method is proposed in this paper, which approximates an object on the ground to a simple bounding solid and works automatically without any prior information about the object. A spherical or cylindrical object determined based on a circularity measure is approximated to a bounding cylinder, while the other general free-shaped objects to a bounding box or a bounding cylinder appropriately. For a general object, its silhouette on the ground is first computed by back-projecting its projected image in image plane onto the ground plane and then a base rectangle on the ground is determined by using the intuition that touched parts of the object on the ground should appear at lower part of the silhouette. The base rectangle is adjusted and extended until a derived bounding box from it can enclose the general object sufficiently. Height of the bounding box is also determined enough to enclose the general object. When the general object looks like a round-shaped object, a bounding cylinder that encloses the bounding box minimally is selected instead of the bounding box. A bounding solid can be utilized to localize a 3D object on the ground and to roughly estimate its volume. Usefulness of our approach is presented with experimental results on real image objects and limitations of our approach are discussed.

Key words: Localization, Object Approximation, 3D Bounding Solid, Monocular Vision

1. INTRODUCTION

3D object reconstruction from images is an attractive and challenging research area in computer vision. It has many applications such as realistic 3D scene creation for entertainment, intelligent surveillance, and autonomous navigation. There is extensive research on 3D object reconstruction from multiple views [1,2]. However, in most cases, only one view of an object or a scene is given. Inference of 3D shapes from a single image is generally under-constrained since depth information of a 3D object is lost through its projection onto the image plane. Thus the 3D object reconstruction from a single view is obviously a harder problem than that from multiple views. Many 3D reconstruction methods [3-8] from a single image have been proposed. In order to resolve the lack of constraints, their applications are confined to piecewise planar objects [5,6], surface of revolution [7], or rectilinear buildings [8] and lots of user interactions are used.

In this paper, we focus on localization of 3D free-shaped objects from a single image without

※ Corresponding Author : Minhwan Kim, Address : (609-735) Department of Computer Engineering, Pusan National University, Busan, Korea, TEL : +82-51-510-2423, FAX : +82-51-517-2431, E-mail : mhkim@pnu.kr
Receipt date : Nov. 11, 2008, Approval date : Dec. 1, 2008
† Dept. of Computer Engineering, Pusan National University
(E-mail : shjung@pnu.kr)
†† Dept. of Computer Engineering, Pusan National University
(E-mail : jangdowon@pnu.kr)
††† Dept. of Computer Engineering, Pusan National University
※ This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program / Institute of Logistics Information Technology)

any prior information about the objects and user interaction. Automatic localization and volume estimation of moving or stationary 3D objects in real 3D world is very useful for intelligent surveillance in autonomous work space. It should be noted that we are interested in only localizing the 3D objects instead of reconstructing them accurately.

An approximation method of an object in a single image to a 3D bounding solid in real 3D world is proposed in this paper, which works automatically when the object is on the ground. The bounding solid such as cylinder or box enclosing a real 3D object can be considered as the simplest and useful representation for estimating its 3D location and volume roughly. The assumption that the object is on the ground is not so strict in real world.

A projected object in input image is first classified automatically into a spheric, cylindric, or general free-shaped object. A bounding cylinder is fitted to a spheric or cylindric object. For a general free-shaped object, its silhouette on the ground is first computed by back-projecting it onto the ground and a base rectangle on the ground is determined by using intuition that touched parts of the object on the ground should appear at lower part of the silhouette. Then a bounding box related to the base rectangle is determined enough to enclose the object. A bounding cylinder is also determined, which encloses the bounding box. The bounding box or the bounding cylinder is selected as the final bounding solid by considering effectiveness and safety of bounding. Usefulness of bounding solids is analyzed through experiments with real 3D objects.

2. APPROACH TO PROBLEM SOLVING

Figure 1 shows general pose of a surveillance camera and relationship between the camera and an object. The 3D work space is represented with the world coordinate system $X_wY_wZ_w$, while the 2D image space with the screen coordinate system

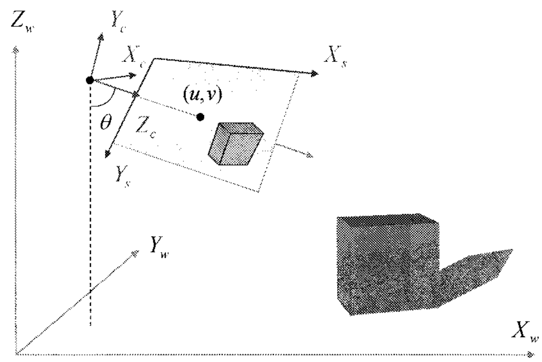


Fig. 1. A camera pose and definition of world, camera, and screen coordinate systems.

X_sY_s. We assume that the 3D object is on the ground (X_wY_w-plane) and the camera is allowed to rotate around the X_c axis (tilting) and Y_c axis (panning) at arbitrary location in the work space. The intrinsic parameters of camera (focal length f, principal point (u, v), x- and y- directional effective pixel size S_x and S_y) and the extrinsic ones (tilt angle θ and height h of the camera) are computed by using the camera calibration method [9,10]. Our goal is to determine a 3D bounding solid enclosing the object and its location in the work space by using information about the projected object in the image space. We assume that area of the projected object is already extracted from an input image.

2.1 Lack of Constraints in a Single Image

Let assume that a rectangular parallelepiped object as shown in Figure 1 is projected on the image plane. Then the projected object may be shown as one of three cases in Figure 2. The boundaries (red bold lines) of the projected object corresponding to the object's boundaries lying on the ground are called *base edges* in this paper. The base edges can be easily determined and the *base shape* (blue rectangle) corresponding to the object's bottom plane on the ground also with the calibrated camera. Then the height of the rectangular parallelepiped object can be computed with coordinate

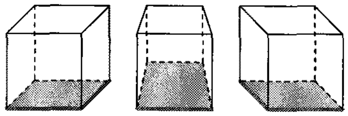


Fig. 2. Three cases of the projected shape for a rectangular parallelepiped object. Red bold lines represent base edges and blue bottom rectangles base shapes.

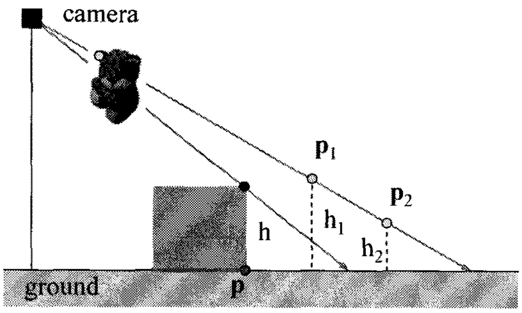


Fig. 3. 3D location ambiguity due to projection.

values of a known point (e.g., the point **p** in Figure 3) on the ground. The coordinate values of **p** can be easily computed by back-projecting the corresponding point **p′** in image space onto the ground, because **p′** is a base edge. However there is no simple way to estimate height of a non-base point because of the 3D location ambiguity due to projection. For example, the yellow floating point of a doll image in Figure 3 is not a base edge point, so its coordinate values in the work space cannot be determined.

2.2 Classification of Object Types

To determine a bounding solid for a 3D object, we first need to estimate the base edges and the base shapes as discussed in previous section. Two types of base shape, circle and rectangle, are suitable for representing the bottom shape of the bounding solid. A circular base is good for a spheric object or a cylindric one, while a rectangular base for a box-like object. For a general free-shaped object, an appropriate base need to be selected because the bottom shape of the general object tends to be irregular complex shape. In this paper, a rectangular base is first estimated for a general free-shaped object and then it will be changed to a circular one if the rectangular base looks like an elongated shape.

A bounding cylinder or a bounding box will be used as a bounding solid which encloses a 3D object on the ground. The former is good for a spheric object or a cylindric one, while the latter for the other objects basically. Therefore, we need a classification method that distinguishes a spheric object and a cylindric one from other types of objects.

Algorithm OBJECT_CLASSIFICATION (a projected object O in image space)

- Compute circularity C from the boundary of O ;
- If $C > T_c$, then determine O as a spheric object and exit ;
- Determine the silhouette of O by back-projecting O onto the ground ;
- Determine a circle most well fitted to lower part of the silhouette ;
- Compute fitness F between lower half of the fitting circle and lower part of the silhouette ;
- Determine width w of O and diameter d of the projected fitting circle in image space;
- If $F > T_f$ and $d > w/2$, then determine O as a cylindric object ;
- else determine O as a general free-shaped object ;

The circularity C of a projected object is defined as $4\pi \times$ area / perimeter^2, whose value is normalized form 0 to 1. A silhouette in work space for the projected object is determined by back-projecting the object onto the ground. Figure 4 shows

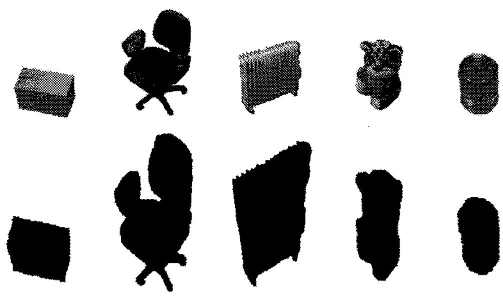


Fig. 4. Examples of silhouette for five different projected objects.

silhouettes for several different projected objects.

The most well fitted circle is determined iteratively by changing the diameter, whose lowest point keeps touching the lowest point of the silhouette. Figure 5 shows examples of determining the circle for two silhouettes. The fitness F is defined as a ratio, (no. of object's boundary pixels overlapped with lower half of the projected fitting circle) / (total no. of pixels corresponding to lower half of the projected fitting circle), as shown in Figure 5. Here we can see that the fitting circle to the water pail is suitable as a circular base while one of the doll is not. Thus another condition for suitability of size, $d > w/2$, is tested.

Fig. 5. Examples of determining the most well fitted circle and computing the fitness and the size suitability.

Through test with several real object images, two thresholds T_c and T_f are selected as 0.9 and 0.7, respectively. Table 1 shows an example of determining object type for three different image objects.

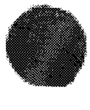
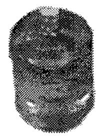
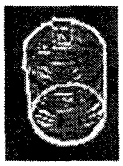
3. CONSTRUCTION OF 3D BOUNDING SOLIDS

Spheric and cylindric objects are approximated to 3D bounding cylinders, while general free-shaped objects to 3D bounding boxes or cylinders.

3.1 Bounding Cylinders for Spheric Objects

A spheric object is touched to the ground at a point $\mathbf{p}$ whose coordinate values are unknown, as shown in Figure 6. However we can compute coordinate values of $\mathbf{p}_1$ and $\mathbf{p}_2$ that are respectively projected to the known points $\mathbf{p}_1'$ and $\mathbf{p}_2'$ in image space. Let lengths of three sides of the triangle $\triangle \mathbf{op}_1\mathbf{p}_2$ be a, b, and c. Then area S of the triangle can be computed from Heron's formula or by using the radius r of the sphere in work space, as shown in (Eq. 1). On the one hand, the angle θ in Figure 6 can be computed by using the tilt angle of camera and the coordinate values of $\mathbf{p}_1'$ in image space. Then the length l can be determined as in (Eq. 2).

Table 1. An example of determining object type.

Object			
Circularity (C)	$0.96 > T_c$	$0.86 < T_c$	$0.67 < T_c$
Fitness (F)	-	 $0.86 > T_f$	 $0.82 > T_f$
Size Suitability	-	$w = 39$, $d = 32$ $d > w/2$	$w = 57$, $d = 22$ $d < w/2$
Object Type	Spheric	Cylindric	General

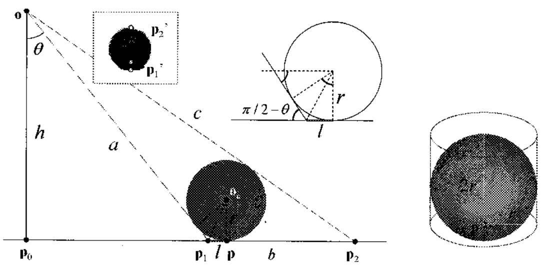


Fig. 6. Construction of a 3D bounding cylinder for a spheric object.

Thus we can determine the point **p** that lies on the line segment $\mathbf{p_1p_2}$. A bounding cylinder for the spheric object can be easily determined by using the radius r, as shown in Figure 6.

$$S = \sqrt{s(s-a)(s-b)(s-c)} = sr,\ s = (a+b+c)/2 \quad \text{(Eq. 1)}$$

$$l = r \times \tan\left(\frac{\pi/2-\theta}{2}\right) \quad \text{(Eq. 2)}$$

3.2 Bounding Cylinders for Cylindric Objects

Sometimes the most well fitted circle determined in the procedure of classifying object types cannot be a base circle of bounding cylinder enclosing a cylindric object, as shown in Figure 7. Radius of the fitting circle needs to be increased, thereby a derived cylinder established on it encloses the cylindric object. Note that the center point of the finally increased circle may be different from original center location of the fitting circle.

In the meantime, height of the derived cylinder on an appropriately adjusted base circle needs to be carefully determined also. An upper circle fitted to upper part of the cylindric object is similarly computed as in the procedure of classifying object

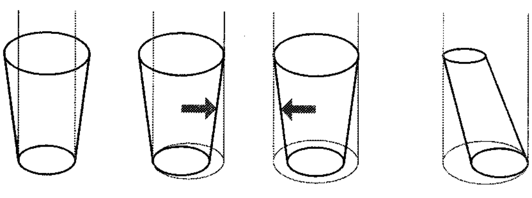


Fig. 7. Procedure of constructing base circles fitted to cylindric objects.

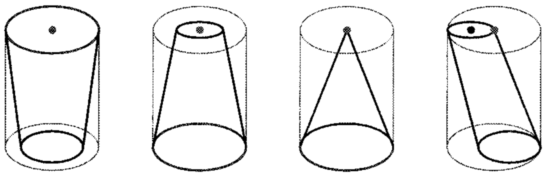


Fig. 8. Examples of determining height of bounding cylinders.

types. If center of the upper fitting circle lies just above center of the base circle in image space, then the height is selected as the center height of the upper fitting circle in work space as shown in Figure 8. If not, we can consider the cylindric object as a skewed object. There is no simple way of determining the height of the skewed cylindric object. In this paper, height of the bounding cylinder for a skewed or not-skewed cylindric object is determined such that center point of the upper fitting circle and one of the upper circle of the bounding cylinder appear at the same row in image space, as shown in Figure 8.

3.3 Bounding Solids for General Free-Shaped Objects

Determining of base circles for spheric or cylindric objects is relatively easy, because a circle is a regular shape. However determination of a base shape for a general free-shaped object is difficult, because bottom shape of the object is irregular. We will first determine base edges in image space, which correspond to points lying on the ground in work space. Then a base rectangle will be determined based on the base edges such that a derived bounding box established on it can enclose the general free-shaped object sufficiently. Finally height of bounding box will be adequately estimated for the object not to penetrate the bounding box.

3.3.1 Estimation of base edges

Figure 9(a) shows three projected general objects and their true base edges. Corresponding base

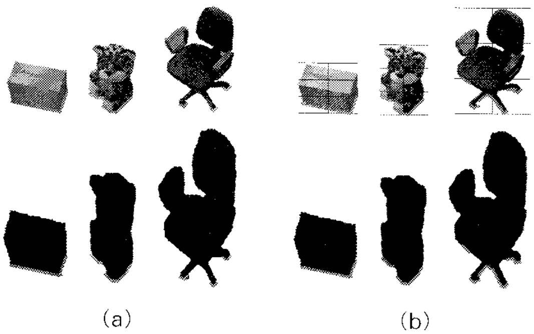


Fig. 9. (a) true base edges for three general objects and corresponding ones on their silhouettes; (b) base edges defined in this paper.

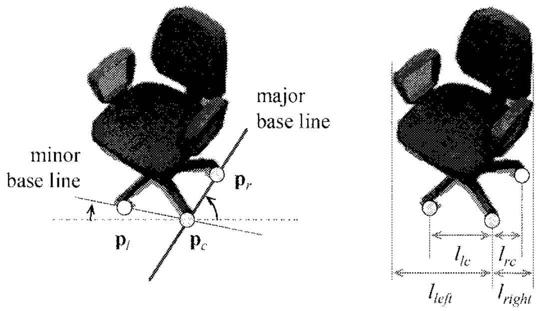


Fig. 10. Determining three base points and two base lines.

edges on their silhouettes in work space are also shown in lower row. Here we can see that the corresponding true base edges on silhouettes remain at their own location in work space. On one hand, the true base edges in image space tend to be at lower part of the projected general objects.

In this paper, base edges are defined as the edges in image space, each of which satisfies the following constraints.

(1) It should lie at lower 1/3 part of the general object.

(2) It should be the lowest point in its own column.

(3) Its downward gradient angle should be in the range of 190° ~ 350°.

The constraints (1) and (2) are intuitively determined through experiments with many real general projected objects. Figure 9(b) shows base edges defined by the above constraints. We can see that almost of the true base edges in Figure 9(a) are included. On the other hand, there are also some wrong ones. However the wrong base edges will not have an effect on what will be done in following procedures.

3.3.2 Determination of base points and base lines

To construct a base rectangle that becomes bottom shape of a bounding box, three base points are first selected in this paler. The center base point $\mathbf{p}_c$ is selected as the lowest base edge in image space, as shown in Figure 10. The right and the left base points, $\mathbf{p}_r$ and $\mathbf{p}_l$, are selected as the first meeting points when the horizontal line passing through the center base point rotates in counter-clockwise and clockwise, respectively, about the center base point. Additionally, the left base point should satisfy the constraint $l_{lc} > l_{left}/2$ and the right one $l_{rc} > l_{right}/2$, for robustness of base point selection. These constraints prevent a base edge near the center base point to be another base point.

A base line is selected as the line passing through the center and the right base point. Another one is the line passing through the center and the left base point. The base line that has longer length between the center base point and its own base point is called the major base line, while the other one the minor base line, as shown in Figure 10.

3.3.3 Determination of base rectangle

Two base lines need to be adjusted to be perpendicular each other, because they become two adjacent sides of the base rectangle. The minor base line is rotated about its own base point until the two corresponding base lines on silhouette become perpendicular each other, as shown in Figure 11.

Then an intial base rectangle can be determined by using the left, right, and adjusted center base points, as in Figure 12(a). However it may be too

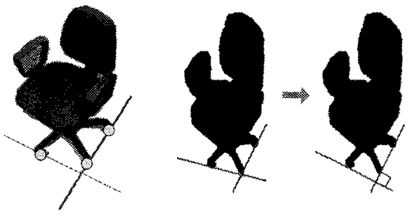


Fig. 11. Making two base lines be perpendicular.

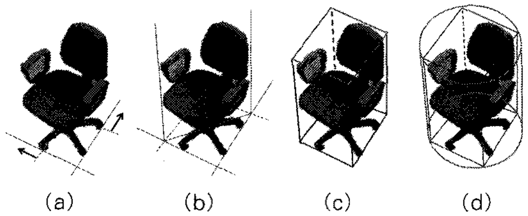


Fig. 12. Construction of a base rectangle and a bounding solid (box or cylinder).

small to enclose the given general object. Thus the initial base rectangle needs to be extended until the derived bounding box from a base rectangle does not encroach upon the object. Figure 12(b) shows the base rectangle that are finally extended by translating the left and right base points along their own base lines as shown in Figure 12(a).

3.3.4 Determination of bounding solid

A bounding box with the base rectangle can be easily determined in work space by pulling down the upper rectangle of the bounding box as long as it does not encroach upon the given general object. In this paper, the upper rectangle is actually pulled down until one of its two rear sides meets the object in image space, as shown in Figure 12(c). A bounding cylinder for the given general object can be also easily determined. The base circle for the bounding cylinder is just the circumcircle of the base rectangle, as in Figure 12(d).

A bounding solid for a 3D object is useful for determining its location and rough volume in work space. We can see that, as location of the object, the center point of the base rectangle or the base circle in Figure 12(d) is much better than any one of the three base points. In the meantime, we know that the bounding box is more compact than the bounding cylinder. However it is possible that the bounding box cannot enclose the given general object sufficiently. On the one hand, the bounding cylinder tends to occupy larger work space than the bounding box, even though the former is more safe than the latter in enclosing the given general object. This conflict may be solved by using the aspect ratio (= length of minor side / that of major side) of the base rectangle. If the aspect ratio is greater than a threshold selected by users, then a bounding cylinder is selected as the bounding solid. For example, in Figure 12(d), the bounding cylinder looks more suitable than the bounding box.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

Figure 13 shows whole sequence of constructing the bounding solid for a real projected chair image

Fig. 13. An example sequence for constructing a bounding solid for a projected chair image.

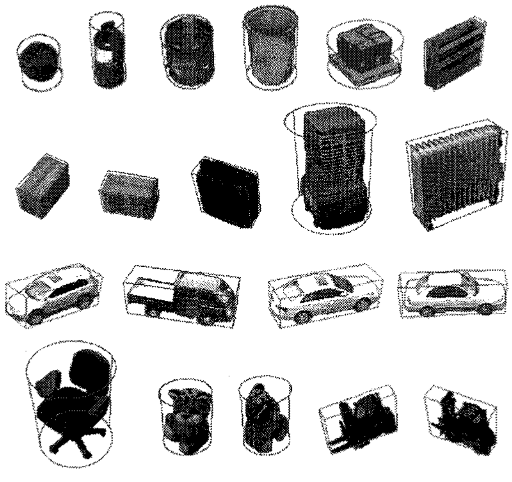


Fig. 14. Experimental results for spheric, cylindric, and general free-shaped objects.

Fig. 15. Real images superimposed with bounding solids.

according to the procedure described in this paper. Figure 14 shows several bounding solids for spheric, cylindric, and general free-shaped objects, which are determined by the proposed method in this paper. The threshold for selecting a bounding solid between a bounding box or cylinder was 0.6 in this experiment. Figure 15 shows four bounding solids superimposed on real images. We can see that the bounding solids are suitable for localizing the objects in Figure 14 and 15.

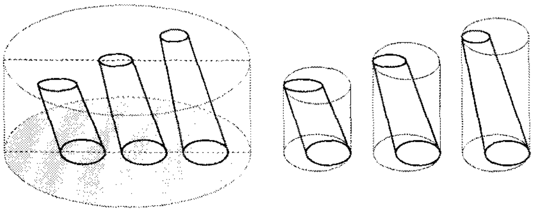


Fig. 16. Undesired results for skewed objects.

However we found that the proposed method was good for non-skewed objects. Three skewed cylindric objects with the same height was shown in Figure 16, but their bounding cylinder are different each other. This undesired results may occur similarly in determining bounding solids for skewed general objects because of limitations caused by using a single image.

Figure 17 shows other problems in the proposed method. The base rectangle tends to be smaller than its actual size, as shown in truck image of Figure 17. Especially, this reducing phenomenon occurs rear sides of a base rectangle. We found that this phenomenon resulted from rounded left or right sides of objects in extending initial base rectangle as described with Figure 12(b). Meanwhile, we can see correct bounding boxes for the paper box and the rectangular parallelepiped in Figure 14,

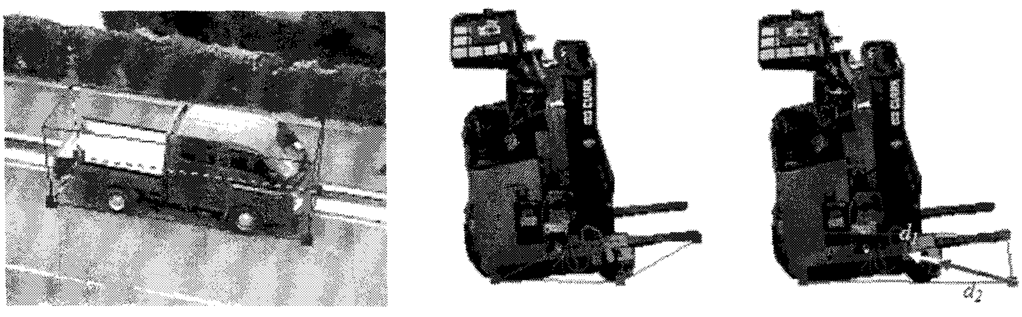


Fig. 17. Examples of side effects occurred in proposed method.

which have perpendicular left and right sides. Another problem is incorrect base rectangles as shown in forklift images of Figure 17.

In further research, solutions for these problems will be studied by using some additional information about target objects such as their symmetric property and major direction of edge lines.

5. CONCLUSIONS

A method for estimating a 3D bounding solid of an object in a single image is proposed in this paper, which works without any prior information about the object or user interaction. The 3D bounding solid can be utilized not only to localization of the object but also to rough volume measurement. We found that the method is useful for approximating non-skewed objects on the ground through experiments with various kinds of real 3D objects. However the method still showed undesired results for skewed objects and very complex shaped objects because of limitations of a single view image. The proposed method is expected to be effectively used for navigation of autonomous vehicles to avoid obstacles and for many other applications requiring of localizing relatively simple objects in real 3D work space. In further research, the method will be improved by using some additional information or knowledge about application environment.

REFERENCES

[1] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, Upper Saddle River, New Jersey, USA, 1998.

[2] S.M. Seitz, et al., "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol.1, pp. 519-528, 2006.

[3] L. Zhang, G.D. Phocion, J.S. Samson, and S.M. Seitz, "Single View Modeling of Free-Form Scenes," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol.1, pp. 990-997, 2001.

[4] D. Jelinek and C.J. Taylor, "Reconstruction of Linearly Parameterized Models from Single Images with a Camera of Unknown Focal Length," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.28, No.4, pp. 767-773, 2001.

[5] Z. Li, J. Liu, and X. Tang, "A Closed-form Solution to 3D Reconstruction of Piecewise Planar Objects from Single Images," *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol.1, pp. 1-6, 2007.

[6] S. Mohan and S. Murali, "Automated 3d Modeling and Rendering from Single View Images," *International Conf. on Computational Intelligence and Multimedia Application*, Vol.4, pp. 476-480, 2007.

[7] C. Colombo, A. Del Bimbo, and F. Pernici, "Metric 3D Reconstruction and Texture Acquisition Surfaces of Revolution from a Single Uncalibrated View," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.27, No.1, pp. 99-114, 2005.

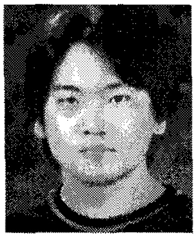
[8] W. Ruisheng and F.P. Ferrie, "Camera Localization and Building Reconstruction from Single Monocular Images," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 1-8, 2008.

[9] Z. Zhang, "A Flexible New Technique for Camera Calibration," *MSR-TR-98-71, Technical Report, Microsoft Research*, 1998.

[10] I. Chen and S. Wang, "An Efficient Approach for the Calibration of Multiple PTZ Cameras," *IEEE Trans. on Automation Science and Engineering*, Vol.4, No.2, pp. 286-293, 2007.

Sunghoon Jung

He received his B.S. and M.S. degrees from Pusan National University, Busan, Korea, in 2006 and 2008, respectively. He is currently a Ph.D. degree student of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system and computer vision.



Dowon Jang

He received his B.S. degree from Silla University, Busan, Korea, in 2008. He is currently a M.S. degree student of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system and computer vision.

Minhwan Kim

He received his B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea, in 1980, 1983, and 1988, respectively. He is currently a professor of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system, multimedia information retrieval, color engineering, and computer vision.