Estimation of a 3D Bounding Box for a Segmented Object Region in a Single Image

Sunghoon JUNG\(^{a)}\), Student Member and Minhwan KIM\(^{b)}\), Nonmember

SUMMARY This paper proposes a novel method for determining a three-dimensional (3D) bounding box to estimate pose (position and orientation) and size of a 3D object corresponding to a segmented object region in an image acquired by a single calibrated camera. The method is designed to work upon an object on the ground and to determine a bounding box aligned to the direction of the object, thereby reducing the number of degrees of freedom in localizing the bounding box to 5 from 9. Observations associated with the structural properties of back-projected object regions on the ground are suggested, which are useful for determining the object points expected to be on the ground. A suitable base is then estimated from the expected on-ground object points by applying to them an assumption of bilateral symmetry. A bounding box with this base is finally constructed by determining its height, such that back-projection of the constructed box onto the ground minimally encloses back-projection of the given object region. Through experiments with some 3D-modelled objects and real objects, we found that a bounding box aligned to the dominant direction estimated from edges with common direction looks natural, and the accuracy of the pose and size is enough for localizing actual on-ground objects in an industrial working space. The proposed method is expected to be used effectively in the fields of smart surveillance and autonomous navigation.

key words: 3D bounding box estimation, object localization, dominant direction of object, back-projection of image region, smart surveillance

1. Introduction

The need for smart visual surveillance systems has recently increased in various fields like security, safety, and facility management, etc. A common objective of these systems is not only to detect, localize, and track objects of interest, but also to automatically interpret and classify their behaviors and activity in an application environment. Therefore, monitoring operators in a traditional passive video surveillance system can be replaced with smart surveillance cameras.

Region localization is defined, in a generic model of smart visual surveillance systems [1], as the transformation of the two-dimensional (2D) image coordinates of a detected region into the corresponding 3D world coordinates, which is considered one of the essential processes needed to meet the requirements for commercial application of smart visual surveillance systems. Usually, a specific point in the region (e.g., the center, the lowest point, or the bottom center of a bounding rectangle) is used to determine 3D location of the region. For example, the representative location estimation method proposed in the system by Collins et al. [2] computes the geolocation of a small object in a far-field scene by intersecting a back-projected viewing ray through the center of the bottom of a region-bounding rectangle with a terrain model of the ground. However, application of the specific point localization technique is limited to small regions because a point can only reasonably represent small regions. In other words, the specific point localization technique fails to handle relatively large image objects in a near- or mid-field scene because the computed location is largely dependent on both camera location and camera angle.

The usefulness of a 3D pose (position and orientation) for an object in tracking processes was discussed by Hu et al. [3], who divided object tracking methods into four major categories: (1) region-based, (2) active contour-based, (3) feature-based, and (4) model-based. Two-dimensional approaches in the first three categories cannot provide 3D pose information, while 3D model-based tracking algorithms in the last category naturally output 3D pose for objects. 3D-pose information is more effective for accurate positioning in 3D space and handling of occlusions than 2D (or 3D) location information. However, 3D model-based tracking algorithms inevitably have some disadvantages, such as the need to construct the models and high computation cost. Thus, their applications are still restricted to the simple specific object models, such as vehicles [4].

On the one hand, 3D size information for objects may be useful in some applications. For example, in autonomous navigation applications, 3D bounding volume information for autonomous-navigation vehicles and their obstacles is very useful for path planning and collision detection. The 3D measurable size information for parts or products from surveillance cameras may also be useful for efficient management of factory facilities.

Although 3D reconstruction methods based on a single image [5], [6] can be one of the ways to obtain 3D size information, their application is restricted to objects with simple structures, such as piecewise planar objects [5] or rectilinear buildings [6], because of the lack of information present in a single image and the high computation cost.

One 3D localization method for an image object [7], which generates a 3D bounding cylinder or a bounding box according to shape of the image object, can also obtain 3D size information. This method reliably constructs bounding cylinders for spherical or cylindrical objects, but it often generates unusual bounding boxes for other types of objects. The fundamental principle of this method is based on the
universal assumption that the lower part of an image object might be on the ground. However, the lower ground portion is estimated empirically and roughly, without considering structural characteristics of the image object; that is, an image point is considered an on-ground point if the image point lies in lower third of the image object and it is the lowest point in its own image column. This empirical estimation sometimes results in incorrect ground segments and unusual bounding boxes. In addition, this method recovers self-occluded backside of 3D objects by applying rotational symmetry to the estimated ground segment, instead of bilateral symmetry, which is usually found in most objects around us. Such recovery of self-occlusion sometimes incorrectly estimates a base for the box, and thus, results in an abnormal bounding box.

A bounding box representation for 3D objects is a simple yet powerful descriptor for their 3D position, orientation, and size. The position of a 3D object can be defined by the center of the base of its bounding box; therefore, the position is robust to changes of relative pose between the viewing camera and the 3D object. The orientation and size of the object can also be reasonably estimated when the bounding box is well aligned to the orientation direction of the object. This 3D information may enable effective classification, identification, and tracking of 3D objects in a surveillance environment. Thus, the bounding box representation can be considered a useful tool in developing 3D-based surveillance algorithms. However, it may be impossible to estimate a bounding box from a single-view image of a 3D object, even though an algorithm for approximating the minimum volume bounding box of a 3D point set was introduced [8]. It is challenging to construct a bounding box for an image object by removing 9 degrees of freedom for positioning a box in 3D space: 3 for position, 3 for orientation, and 3 for size.

In this paper, a novel method for automatically constructing a bounding box for an image object is proposed. The method first determines the base of the box from a set of points expected to be on the ground and then determines the height of the box. Dominant directions of the image object are detected and used to estimate a natural orientation for the bounding box, which appear frequently in man-made object images [9], [10]. Observations associated with structural characteristics of back-projected image objects on the ground are suggested, and their usefulness for determining the expected on-ground points is verified. The tendency for many man-made and natural objects to be symmetric [11], [12] is assumed and exploited to recover unobservable information of the self-occluded backside of objects. The proposed method is designed to work well on objects lying on the ground in order to reduce degrees of freedom for positioning a box in 3D space.

The remainder of this paper is organized as follows. A definition of the given problem in this paper is introduced in Sect. 2, and a broad outline for solving this problem is discussed. Domain transformation from image space to ground space is introduced in Sect. 3, and some useful properties in

ground domain are exploited and discussed. A technique for determining dominant directions in an image object is also introduced and discussed in Sect. 3. Some useful observations for estimating on-ground points are suggested and their usefulness is verified in Sect. 4. A method for determining a bounding box based on symmetry is described in Sect. 5. Usefulness of the proposed method is confirmed by experiments on real objects and 3D-modelled objects in Sect. 6, and a conclusion follows in Sect. 7.

2. General Overview for Problem Solving

A general relative pose between a surveillance camera and a 3D object on the ground in some working space is shown in Fig. 1. The camera is calibrated by using Zhang’s calibration method [13] with a calibration pattern on the working floor. The known points on the calibration pattern can be written as coordinate vectors of the form \((x, y, 0)\), thereby the ground (that is, the working floor) is represented as the \(x_0y_0z_0\) plane in world coordinate system \((x_0, y_0, z_0)\).

Our objective is to determine a bounding box for the 3D object on the ground from a region of the object on the image plane, called an image object in this paper.

Generally, a box in 3D world space has 9 degrees of freedom (doF): 3 for position, 3 for orientation, and 3 for size. The doF for position can be represented as a point \((x, y, z)\) and the doF for orientation as a set of rotations \(\{r_x, r_y, r_z\}\) about the dimension axis. The doF for size can be written as a set \(\{w, d, h\}\), where \(w, d,\) and \(h\) represent width, depth, and height of the box, respectively.

To reduce the degrees of freedom, some reasonable constraints or knowledge has been used. Information from multiple images or prior knowledge about target objects is widely and generally available. However, we cannot use the information because only a single image object is given here. Thus, we set up the practical and reasonable constraint that a 3D object is on the ground. Then the bounding box for

![Fig. 1 General relative pose between a surveillance camera installed at \(x_c\) and a 3D object on the ground (the \(x_0y_0\) plane); \(x_p\) is the vertical projection of \(x_c\) onto the ground. The shaded region on the ground, called a virtual shadow, can be considered as the back-projection of the image of the 3D object on the image plane, called an image object. The contour generator (bold line segments) on the 3D object is projected to the (bold) boundary of the image object, called its apparent contour. The colored line segments represent a bounding box for the 3D object.](image)
the object is also on the ground and it has 6 dof, as shown in Fig. 2, since 3 dof (1 for position and 2 for orientation) can be removed [4].

The remaining degree of freedom for orientation can be removed by using information about the dominant directions that can be found in most man-made objects [9], [10]. For example, in a wooden bench, long pieces of wooden boards may show a horizontal dominant direction, whereas wooden legs have a vertical dominant direction. We can imagine that non-vertical dominant directions of an object are usually related to orientation of the object, and a bounding box aligned to them looks natural.

On the one hand, estimation of on-ground points in an image object is very helpful for determining two variables corresponding to the remaining dof for position. The on-ground points usually lie at a lower part of the image object [7], and they have some common structural characteristics. Observations associated with the structural characteristics will be proposed later, and their usefulness for determining the expected on-ground points will be verified in this paper.

The three dof for size still remain. Fortunately, there is a general symmetry property whereby many man-made and natural objects are symmetric [11], [12]. In particular, the most common symmetry in our environment is bilateral symmetry [14]. In this paper, bilateral symmetry will be used to recover information about self-occluded backsides of objects, thereby resulting in removal of two dof corresponding to the size of the base of the bounding box. The last degree of freedom for size related to the height of the bounding box will be removed by adjusting the top of the bounding box so as to minimally bound a given image object.

3. Useful Properties in the Ground Domain

Let us consider a back-projecting ray from the origin of a surveillance camera to each boundary point of the image object of a 3D object, as shown in Fig. 1. Each back-projecting ray meets at a point on the surface of the 3D object, called a tangent point, and then makes an intersection point on the ground. Then a set of the boundary points of the image object, called an apparent contour, is the projection of a set of the tangent points onto the image plane, called a contour generator [15]. A set of the intersection points on the ground is called a shadow contour in this paper because it looks like the contour of the shadow of the 3D object when the camera is considered a light source. A set of intersection points for all the points of an image object is called a virtual shadow by similar reasoning.

The contour generator of the bounding box in Fig. 1 consists of three types of contour (base contour, side contour, and top contour) which are represented by red, blue, and green solid line segments, respectively. The contour generator is mapped to a unique apparent contour on the image plane and a unique shadow contour on the ground plane. The former encloses the image object, and the latter the virtual shadow, as shown in Fig. 3. Thus a bounding box for an image object can be constructed in the ground domain instead of in the image domain, such that a virtual shadow of the bounding box encloses a virtual shadow of the image object. A task can be accomplished more efficiently in the ground domain than in the image domain, because there are many useful properties in the ground domain.

3.1 Domain Transformation

The transformation between a scene plane and an image plane is a projective transformation, usually called a homography. Thus, the back-projection from an image plane to the ground plane is also a homography. Let a 3 × 4 matrix \( \mathbf{P} \) be a camera matrix [15]. Then a 2D point \( \mathbf{x} \) in an image co-
ordinate system that is the projection of a 3D ground point \( \tilde{s} = [x y 0 1]^T \) can be written as a homography of a 2D point \( \tilde{x}_c = [x y 1]^T \) in a ground coordinate system, as shown in Eq. (1). All the points are represented in homogenous coordinate form, and the ground coordinate system is defined by the \( x_w \)-axis and the \( y_w \)-axis in the world coordinate system. The homography \( H \) can be represented as \( [p_1 \ p_2 \ p_3 \ p_4] \), where \( p_i \) represents the \( i \)-th column vector of \( \mathbf{P} \). Then the back-projection can be performed by using the inverse transformation \( H^{-1} \).

\[
\tilde{x}_c = \mathbf{P} \tilde{s} = [p_1 \ p_2 \ p_3 \ p_4][x \ y \ 0 \ 1]^T = [p_1 \ p_2 \ p_4][x \ y \ 1]^T = \mathbf{H}\tilde{x}_c
\]

(1)

The camera matrix \( \mathbf{P} \), determined through camera calibration [13], can be represented as \( \mathbf{K}\mathbf{R}[\mathbf{I} - \mathbf{x}_c] \), where \( \mathbf{K} \) and \( \mathbf{R} \) are the camera calibration matrix and the camera orientation matrix, respectively, and \( \mathbf{x}_c \) represents the camera position, as shown in Fig. 1. Thus \( \mathbf{H} \) can easily be determined after calibrating a given camera. Therefore, a shadow contour can be computed from an apparent contour of a given image object by using back-projection \( H^{-1} \). Figure 4 shows the result of a domain transformation for an image object of a small table.

3.2 Properties in the Ground Domain

There are some useful properties in dealing with shadow contours in the ground domain, compared to dealing with image objects or their apparent contours in the image domain.

(Property 1) A shadow contour is uniquely determined dependent only on the position of a given camera.

A virtual shadow of an object can be uniquely determined from an image for the object when the surveillance camera and the object are fixed in 3D world space, as shown in Fig. 1. If the orientation of the camera is changed, then the orientation of the image plane is changed and thus the image of the object is also changed. However, the virtual shadow is not changed. On the other hand, if the position of the camera is changed, then the virtual shadow is also changed. Thus we can say that a virtual shadow and its shadow contour are uniquely determined dependent only on the position of the camera.

(Property 2) A vertical line in 3D world space is back-projected to a ground line that passes through the vertically projected point of a camera origin onto the ground.

The vertically projected point of camera origin \( \mathbf{x}_c = (x_c, y_c, z_c) \) onto the ground, \( (\mathbf{x}_p \text{ in Fig. 1}) \), can be represented as \( (x_c, y_c, 0) \), which is called the projected origin of the camera. On the one hand, as shown in Eq. (2), a vanishing point \( \tilde{v} \) of homogenous form for lines with direction \( \mathbf{d} \) in 3D world space can be written as \( \mathbf{K}\mathbf{R}\mathbf{d} \) in the image coordinate system. Similarly, a 3D point \( (x, y, 0) \) on the ground is projected to a point \( \tilde{x} \) of homogenous form in the image coordinate system, as shown in Eq. (3).

\[
\tilde{v} = \mathbf{P} \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} = \mathbf{K}\mathbf{R}[\mathbf{I} - \mathbf{x}_c] \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} = \mathbf{K}\mathbf{R}\mathbf{d}
\]

(2)

\[
\tilde{x} = \mathbf{K}\mathbf{R}[\mathbf{I} - \mathbf{x}_c][x \ y \ 0 \ 1]^T = \mathbf{K}\mathbf{R}[x - x_c \ y - y_c \ z_c]^T
\]

(3)

According to Eq. (3), the projected origin of the camera, \( (x_c, y_c, 0) \), is projected to \( \mathbf{K}\mathbf{R}[00 - z_c]^T \) in the image coordinate system. From Eq. (2), we can see that the point \( \mathbf{K}\mathbf{R}[00 - z_c]^T \) is equal to the vanishing point \( \tilde{v} \) for lines with a vertical direction \( \mathbf{d} = (0, 0, -z_c) \). Thus, this vanishing point \( \tilde{v} \) is back-projected to the projected origin of the camera, \( \mathbf{x}_p \), and the back-projection of an imaged line that passes \( \tilde{v} \) also passes \( \mathbf{x}_p \). Therefore, a vertical line in 3D world space is back-projected to a ground line that passes \( \mathbf{x}_p \).

All the shadow contour segments corresponding to vertical lines in 3D world space pass \( \mathbf{x}_p \). In Fig. 3, two solid blue line segments pass \( \mathbf{x}_p \) because they correspond to vertical sides of the bounding box in 3D world space. Conversely, this property enables us to guess that a line passing \( \mathbf{x}_p \) in the ground plane, called a support line, may correspond to a vertical line or plane in 3D world space. For example, in Fig. 4(b), the line segment corresponding to a leg seems to be a vertical line in 3D world space because it is overlapped by a blue-dotted support line. Notice also that a vertical plane intersected at a support line with the ground plane is back-projected to the support line itself. We can imagine in Fig. 3 that the plane passing both \( \mathbf{x}_c \) and a side contour of the bounding box in 3D world space might be back-projected to the support line passing \( \mathbf{x}_p \) and the corresponding shadow contour.

(Property 3) Parallel lines in 3D world space are back-projected to parallel ground lines if each line is also parallel to the ground plane.

It is evident that a line \( l_1 \) parallel with the ground plane is back-projected to a ground line parallel to line \( l_1 \). Then, the back-projected ground line is also parallel to another line, \( l_2 \), that is parallel to line \( l_1 \) in 3D world space. Sim-

![Fig. 4 An example of domain transformation for an image object: (a) a small 3D-modelled table, and (b) a virtual shadow of the table and its (bold) convex contour with three (color-dotted) support lines.](image-url)
ilarly, the back-projected ground line of line $l_2$ is parallel to line $l_1$. We can see that all four lines are parallel to one another. Therefore, two parallel lines in 3D world space are back-projected to two parallel ground lines if each line in 3D world space is parallel to the ground plane. Notice that back-projection of any two parallel lines in 3D world space forms a vanishing point on the ground plane if each line is not parallel to the ground plane. We can see that Property 2 deals with special parallel lines, that is, vertical lines to the ground plane.

(Property 4) The convex hull of a shadow contour for an object is enclosed by a shadow contour for the bounding box of the object.

A set of points is defined as convex if it contains all the line segments connecting each pair of its points. Then the convex hull of a given set $S$ may be defined as the unique minimal convex set containing $S$. A bounding box of an object encloses the 3D convex hull of the object as well as the object itself. Thus, under a given camera installation, it is evident that the virtual shadow for the bounding box contains the virtual shadow for the 3D convex hull as well as one for the object. It is also evident that the former, the virtual shadow for the 3D convex hull, is a convex set and minimally contains the latter. Thus, the former is the convex hull of the latter. Consequently, the shadow contour for the bounding box of the object encloses the convex hull of the virtual shadow for the object. From this consequence, we can see that only the boundary of this convex hull, called a convex contour, affects estimation of the bounding box. A convex contour for an object is usually much simpler than a shadow contour for the object, as shown in Fig. 4(b).

(Property 5) Only visible convex vertices from a projected origin of the camera can be object points on the ground.

Let us consider a convex contour as a convex polygon. Then, a convex contour may be defined as a sequence of vertices of the convex polygon. A vertex of a convex contour, called a convex vertex, is defined as visible when a ray from a projected origin of camera, $x_c$, toward the convex vertex meets the vertex itself first. In other words, a convex vertex is invisible if there are some internal points of the convex contour between $x_c$ and the convex vertex on the support line passing these two points. Under the reasonable assumption that a camera is not installed with a steep downward orientation, a 3D object usually looks self-occluded, and an invisible convex vertex corresponds to an object point not on the ground, even though the object point is directly visible from the camera center $x_c$, as shown in Figs. 1 and 3. On the other hand, visible convex vertices may be object points on the ground under the assumption of normal camera installation. A convex vertex is referred to as on-ground if it is actually an object point on the ground, otherwise it is referred to as off-ground. Notice that not all the visible convex points are on-ground. For example, in Fig. 4, the visible convex vertex on the left red support line is not on-ground.

A similar observation was proposed by Jung et al. [7], where they argued that a lower part of an image object may correspond to object points on the ground. This empirical observation leads to a rough or unwarranted assertion that an image point corresponds to an object point on the ground if the image point lies in the lower third of the image object and it is the lowest point in its own image column. In this paper, the empirical observation can take forms such that only visible convex vertices may be object points on the ground under the assumption of normal camera installation. We can see that the lower part of an image object to Jung et al. [7] is refined, in this paper, to the lower part of the convex contour for the image object. Notice that the lower part of the convex contour is confined by the two support lines that touch the convex contour at a point or a line segment. These support lines are shown as red-dotted lines in Figs. 3 and 4; one is called a left tangent support line, and the other, a right tangent support line. A convex vertex on a tangent support line is called a tangent vertex in this paper.

3.3 Determination of Dominant Directions in the Ground Domain

Dominant directions are frequently found in man-made object images [9, 10]. In particular, a dominant direction parallel with the ground plane is very useful because it is usually related to orientation of the object. We have seen that a bounding box aligned to a dominant direction looks very natural.

A dominant direction is usually represented in an object as multiple lines in this direction. Thus a dominant direction can be determined by finding those lines with the same orientation. However, in a given image object, those lines are represented as lines with different orientations to each other because of perspective projection. We know that those lines form a vanishing point on the image plane. Thus dominant directions in an image object can be determined by detecting the vanishing points.

On the one hand, here we want to identify only the dominant directions parallel to the ground plane. We know from Property 3 that only the parallel lines with a direction parallel to the ground plane are represented as parallel lines in the ground domain. In other words, the other parallel lines in 3D world space are shown as non-parallel lines in the ground domain. Thus, a useful dominant direction for determining natural orientation of an object can easily be determined by finding lines with a common direction in the ground domain.

In this paper, some worthwhile edges for forming lines in the image domain are first selected by using the Hough transformation [16, 17], and they are represented in the ground domain through the domain transformation. Then, dominant directions are determined by selecting peaks in the directional histogram where the number of edges in each line is accumulated at the bin corresponding to orientation (or slope) of the line, as shown in Fig. 5. In this approach,
the vertical direction to the ground plane is often selected as a dominant direction, because lines closely similar to the vertical direction are identified frequently as lines with the same orientation by the Hough transformation, and these lines form a peak in the directional histogram. Fortunately, the direction vertical to the ground plane can easily be identified and excluded by comparing orientation of each peak with the average orientation of the support lines for visible convex vertices.

4. Estimation of On-Ground Point

The core task for constructing a bounding box for a target object is estimation of a proper base from its convex contour, because the estimated base is closely related to location and orientation of the target object. The most important thing in estimating a proper base is estimation of on-ground points among convex vertices of the convex contour, because the base will be estimated from only the on-ground points. From Property 5 in Sect. 3.2, we know that only the visible convex vertices can be on-ground points. Those visible vertices form a sub-sequence of the convex contour, called a candidate convex sequence.

To develop a method for assigning a proper attribute (on-ground or off-ground) to each visible convex vertex, we observed and analyzed structural characteristics of convex vertices in various types of objects. Here, some observations for assigning attributes to visible convex vertices are introduced, and their usefulness is evaluated through testing with various 3D-modelled objects. Finally, a linear ordering method is suggested to ease conflicts among the observations.

4.1 Observations

There are two types of observations. One is directly applicable to a convex contour, and the other is related to dominant directions. A convex segment with a dominant direction in a candidate convex sequence is called a dominant segment, which will be used effectively in two observations for determining on-ground points.

On the one hand, we know that one of the sides of a minimum bounding rectangle should be collinear with one of the sides of the convex hull of a shape [18], and a method called rotating calipers [19] obtains a minimum bounding rectangle for a set of points in linear time. The rotating calipers consist of two parallel line sets perpendicular to each other. The rotating calipers method makes a line from a parallel line set collinear to a convex segment of a convex hull of the point set, and then constructs a bounding rectangle enclosing the convex hull. Among the bounding rectangles for each convex segment, the bounding rectangle with the minimum area is selected as the minimum bounding rectangle. However, this method is not directly applicable to our problem of determining a base with the minimum area from a set of on-ground points, because on-ground base points that are self-occluded on the backside of a base are not available.

In this paper, a try square is used to find a natural bounding box with minimum base area, which consists of two line segments forming a right angle at a reference point. Here one of the line segments is called a base line, and the other, a subsidiary line. A try square for each dominant segment will be oriented such that its base line is collinear to the dominant segment, and its subsidiary line touches a convex vertex or a convex segment. Later, among bases derived from each try square, the base with the minimum area will be used to construct a natural bounding box.

(Observation 1) If a convex vertex is not visible, then it is off-ground.

From Property 5 in Sect. 3.2, only the visible convex vertices can be on-ground vertices. In other words, invisible convex vertices are off-ground. On the one hand, we know that a tangent support line can touch a convex segment. That is, there may be two tangent vertices on a tangent support line. Then, the tangent vertex nearer to the projected origin of the camera is visible, while the other is invisible. According to this observation, the invisible tangent vertex is off-ground.

(Observation 2) If two adjacent visible convex vertices form a convex segment with a slope nearly equal to that of a tangent support line, then the vertex farther from the projected origin of the camera is off-ground.

If a convex segment consisting of two adjacent visible convex vertices has a slope similar to that of a tangent support line, then it is probable that the convex segment corresponds to a vertical line segment in 3D world space. Thus, it is natural that its upper vertex is set off-ground. Such a vertical-like segment occurs usually near tangent vertices. Sometimes, a visible tangent vertex is set off-ground by this observation.

(Observation 3) If each tangent support line has only one tangent vertex that is adjacent to a concave shadow vertex, then the tangent vertex is off-ground. In the case of two tangent vertices on each tangent support line, visible tangent vertices are off-ground if they have an adjacent concave shadow vertex.

Through observation on several objects with a wider
body than their base, we found that visible tangent vertices were probably off-ground when there are concavities around both of the visible tangent vertices in shadow contours. In such a case, we also found that the visible tangent vertices had the same structure of tangency. That is, the left and right tangent support lines meet at one tangent vertex with a convex contour or both of the tangent support lines at two tangent vertices. For example, the wider top of a small 3D-modelled table in the left image of Fig. 6 meets at one tangent vertex with each tangent support line, and there are concavities below both of the tangent vertices in the virtual shadow of this table. These two visible tangent vertices are off-ground, according to this observation. If the wider top of the table is very thick, then there may be two tangent vertices on each tangent support line. Then, the lower visible tangent vertex on each tangent support may be set off-ground according to this observation.

(Observation 4) If a visible convex vertex has an adjacent shadow vertex on its support line, then it is on-ground.

From Property 2, a shadow segment lying on the support line of a visible convex vertex probably corresponds to a vertical line segment standing on the ground, as shown in the left image of Fig. 6. Hence the visible convex vertex can be set on-ground. However, we cannot apply this observation to visible tangent vertices because, as discussed in Observation 3, they may not be on the ground in the case of objects with a wider body than their base.

(Observation 5) If a candidate convex sequence is a subsequence of a shadow contour, and there are two tangent vertices on each tangent support line, then the lower visible tangent vertices are on-ground.

When a 3D object is a prismatic object or a cylinder, its visible tangent vertices are on the ground. Such an object can be verified from the following two conditions. First, its candidate convex sequence should be a part of its shadow contour, because the candidate convex sequence corresponds to a part of the convex base of the object. Second, each tangent support line should overlap a convex segment of its convex contour because the object has vertical sides. Thus, if these two conditions are satisfied, the lower vertex of the convex segment (that is, the lower visible tangent vertex) can be set on-ground.

(Observation 6) If a convex vertex is on an oriented try square, then it is on-ground.

Once a try square is oriented, at least three convex vertices are on the try square. Since the try square is ultimately understood to be part of the base of the bounding box on the ground, the convex vertices on the try square can be marked as on-ground simultaneously.

(Observation 7) Convex vertices of dominant segments in a candidate convex sequence are on-ground.

We found that a convex segment at a lower part of a convex contour for a given object is probably on the ground when it is aligned to one of the dominant directions. This observation is especially useful for an object with a non-square base, because it allows two or more convex segments at non-right angles to be understood as on-ground simultaneously.

Figure 6 shows examples of applying the above observations to two 3D-modelled objects. The observation applied to each convex vertex is denoted by its number near the vertex. A vertex to which multiple observations apply is denoted by numbers separated with commas, while a vertex with no attribute is discriminated with the number 0. Meanwhile, the aligned try square used to apply Observation 6 is also displayed, with the basis line indicated by a bold line. We can see that convex vertices corresponding to object points on the ground are properly attributed by on-ground observations, (Observations 4–7). On the other hand, we can also see that both on-ground and off-ground attributes are assigned to a vertex in the right image of Fig. 6. Such conflicts of assignment will be resolved in Sect. 4.3.

4.2 Usefulness of Observations

A set of 3D-modelled objects from the web [20], which is a collection of 150 image objects rendered from 70 different modelled objects, is used to test usefulness of the proposed observations. To 2,514 convex vertices generated from the image objects, a total of 2,664 assignments are executed by the proposed observations. Thus, we guess that some convex vertices are assigned more than two observations, as shown in Fig. 6. On the one hand, an average 15.4% of the convex vertices in a convex contour are not attributed.

Table 1 shows the number of right and wrong assignments for each observation and their accuracy. All the observations show a high accuracy above 91.7%. Observation 1, with an accuracy of 100%, says that all the image objects are properly captured under normal camera installation. We can see that Observations 2 to 4 and Observation 7 are practical and very useful, because they have many assignments.
and show a very high accuracy. Meanwhile, Observation 5 has relatively few assignments, but it is also a useful observation applicable to objects with side vertical lines, such as prismatic or cylindrical objects. Observation 6 can be applied multiple times to a candidate convex sequence when there are multiple dominant segments. On the other hand, the very high accuracy of Observations 6 and 7 says that information about dominant directions in an image object is very useful for determining ground parts of the image object.

4.3 Linear Ordering of Observations

Sometimes, two different attributes, on-ground and off-ground, are assigned simultaneously to a convex vertex. For example, in the right image of Fig. 6, the right visible tangent vertex satisfies all the conditions for Observations 3, 6, and 7, and thus it is assigned on-ground and off-ground simultaneously. Actually, the tangent vertex corresponds to an on-ground point in 3D world space, and thus, it should be assigned on-ground. However, Observation 3 assigns the wrong attribute (off-ground) to the tangent vertex. Here, we try to resolve such conflicts of assignment by restricting assignment of attribute only to convex vertices with no attribute yet. For example, Observation 3 is restricted from assigning off-ground to the right visible tangent vertex if the vertex is already assigned on-ground by Observations 6 or 7. Such a resolution forces the proposed observations to be linearly (or totally) ordered [21].

A conflict of assignment between observation \( x \) and observation \( y \) can be represented as an ordered pair \((x, y)\) or \((y, x)\). When observation \( x \) assigns the wrong attribute more frequently than observation \( y \), the conflict of assignment is represented as \((x, y)\), and we say that observation \( x \) is inferior to observation \( y \). Then an inferior relation \( R \) on a set of the proposed observations can be defined by \((x, y) \in R\) if observation \( x \) and observation \( y \) make one or more conflicts of assignment, and observation \( x \) is inferior to observation \( y \). We can see that the inferior relation \( R \) is anti-symmetric, because both \((x, y)\) and \((y, x)\) cannot be in \( R \). If \( R \) is acyclic, its transitive reflexive closure is a partial order, and a linear order can be obtained from the partial order by using the topological sorting algorithm [21]. There may be many different linear orders that do not violate the order relationship in the partial order. We can find the most accurate linear order among them by selecting the observation with the highest accuracy first during topological sorting, unless the order relationship is violated.

Table 2 shows a conflict matrix \( C \) of observations for the 3D-modelled objects. An \((i, j)\) entry \( c_{ij} \) is the number of times that observation \( i \) assigns the wrong attribute to a convex vertex, but observation \( j \) assigns the correct one to the vertex. For example, we can see from \( c_{24} \) that Observation 2 assigns off-ground to two different on-ground convex vertices, but Observation 4 correctly assigns on-ground to them. There is no conflict of assignment among Observations 1 to 3, because they assign the same attribute: off-ground. Similarly, Observations 4 to 7, assigning on-ground, do not create conflicts of assignment among them. Table 2 shows seven conflicts of assignment from five conflict cases among 2,664 assignments.

The inferior relation \( R \) on the set of the observations can be written as \( R = \{(2, 4), (3, 6), (3, 7), (5, 2)\} \). Here notice that \((4, 2)\) is not in \( R \) because Observation 2 assigns wrong attributes more frequently than Observation 4. The relation \( R \) is anti-symmetric and has no cycle. Thus, the transitive reflexive closure of \( R \) is a partial order [21]. Figure 7 shows a Hasse diagram for the partial order, where a node and its real number represent the corresponding observation and its accuracy, respectively. The most accurate linear order by using the topological sorting algorithm [21], based on the strategy of selecting the observation with highest accuracy first, can be written as \( 1 \rightarrow 6 \rightarrow 7 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 5 \). If assignment of attribute is executed in this linear order and restricted only to convex vertices with no attribute yet, then four conflicting cases (in bold in Table 2) are resolved, and therefore, the bold numbers in Table 2 are reset to 0. Then, the total number of wrong assignments in Table 1 is reduced from 18 to 12, and thus, it is just 0.50% of 2,664 assignments.

5. Determination of a 3D Bounding Box

On-ground points for each dominant segment can be de-
Fig. 8  Flowchart for determining a 3D bounding box from an input image object.

determined by using the observations proposed in Sect. 4.1. When there is no dominant segment, we make relatively long convex segments play the role of dominant segments. That is, on-ground points for each relatively long convex segment are determined by using the proposed observations. When there is no dominant segment, a try square is oriented such that its base line is collinear to a relatively long convex segment, and then Observation 6 is applied in order to determine on-ground points. Notice that a try square may be collinearly aligned with a dominant segment or a relatively long convex segment. Such a segment is called a try square-aligned segment in this paper.

When a set of on-ground points for a dominant segment or a relatively long convex segment is given, the base of a bounding box can be estimated under the assumption of the general symmetry property [11], [12], [14]. Among the estimated bases for every dominant segment or every relatively long convex segment, a base with the minimum area is selected as the desirable one. Finally, the height of a bounding box with the desirable base is determined, such that the shadow contour for the bounding box encloses the virtual shadow of the given image object. Figure 8 shows a flowchart for determining a 3D bounding box from an input image object.

5.1 Determination of On-Ground Range

Not all convex vertices are assigned an attribute by the linear ordered observations. We saw that an average 15.4% of convex vertices in a convex contour were not attributed in the test for the 3D-modelled objects. Notice that all invisible convex vertices are assigned as off-ground by Observation 1. Thus, almost all unattributed convex vertices in a convex contour are in its candidate convex sequence.

Here, we assume that on-ground and off-ground convex vertices do not appear alternately in a candidate convex sequence when a camera is normally installed (that is, not installed with a steep downward orientation). Then, the leftmost and the rightmost on-ground convex vertices determine a sub-sequence of the candidate convex sequence, called an on-ground range, which may be considered a non-occluded part of the ground section of an object. Figure 9 (a) shows a convex contour with observation numbers applied in the restricted linear order assignment of attributes. We can guess a reasonable on-ground range from the on-ground convex vertices labeled with blue numbers.

5.2 Test of Symmetry

An on-ground range for an object provides information for the non-occluded part of the ground section of the object, but it usually cannot cover more than half of the ground section. Thus, an on-ground range is not enough, by itself, for estimating the whole ground section of an object. Fortunately, there is the general symmetry property, whereby many man-made and natural objects are symmetric [11], [12]. This general symmetry property allows an estimation of the self-occluded backside of the ground section. The rotational symmetry property found frequently around us can be used [7]. However, this property usually results in a smaller base for the bounding box of an object than the original ground section of the object. On the other hand, the bilateral symmetry property found most commonly in our environment [14] lets us estimate a base of almost equal size to that of the original ground section. Thus, in this paper, the bilateral symmetry property is mostly used to recover unobservable information of self-occluded backsides of objects. The rotational symmetry property may also be utilized, but only when the bilateral symmetry property cannot be used.
First of all, bilaterally symmetrizable convex segments are determined by using a reflect-and-test procedure, called a bilateral symmetry test in this paper. In a bilateral symmetry test on a convex segment in a given on-ground range for an image object, all the convex vertices are reflected about the bisector of the convex segment. If all the reflected points fall inside the virtual shadow of the image object, then the convex segment is considered bilaterally symmetrizable.

The bilateral symmetry test is executed on all dominant segments and all relatively long non-dominant segments with a length over one-quarter of the longest segment in the on-ground range.

Bilaterally symmetrizable convex segments in an on-ground range are sequentially applied to generate the points expected to be on the ground. A bilaterally symmetrizable convex segment is accepted and used in estimating a base in cases where the bilaterally reflected points for all the generated points from the previously applied segments fall inside the virtual shadow currently considered. Thus, acceptance of a bilaterally symmetrizable convex segment depends on the order for applying bilateral symmetry. If a bilaterally symmetrizable convex segment is not accepted, then applying bilateral symmetry to the remaining segments in the application order list is not executed, and the set of all generated points from the previously applied segments is considered the final set of generated points.

On the one hand, bilaterally symmetrizable convex segments in an on-ground range can be partitioned into three disjoint sets: (1) a set for a try square-aligned segment and/or its orthogonal convex segment, (2) a set of dominant segments, and (3) a set of relatively long non-dominant convex segments. The segments in the first set play the most important role in estimating a base with the same direction as that of the try square-aligned segment, and thus, their bilateral symmetry needs to be applied first.

The segments in the second set are good candidates to be applied next, because they usually provide more useful information for estimating a meaningful base than those in the third set. The relatively long convex segments in the third set may be used when the segments in the first and the second sets do not generate enough points to estimate a meaningful base.

Now, all the bilaterally symmetrizable convex segments can be linearly ordered if the segments in each set are sorted separately. It seems evident that the try square-aligned segment in the first set precedes its orthogonal convex segment in order. However, it is not easy to find a solid foundation for sorting the segments in the second and third sets. Here, for each bilaterally symmetrizable convex segment in a set, we first compute a convex hull [22] enclosing the convex vertices in the on-ground range and their reflected points about the bisector of the convex segment, and we then determine the minimum bounding rectangle enclosing this convex hull by using the rotating calipers method [19]. Then, the difference in area between the minimum bounding rectangle and the convex hull can be computed for each bilaterally symmetrizable convex segment, which is called a non-occupied area. We expect that a good bilaterally symmetrizable convex segment has a small non-occupied area. Thus, the segments in the set are sorted with respect to their non-occupied areas in ascending order.

On the other hand, if there is no bilaterally symmetrizable convex segment, or no acceptable one, the rotational symmetry property may be used. A set of the points expected to be on the ground can be obtained by rotationally reflecting all the convex vertices in a given on-ground range about the center point of the rightmost and the leftmost convex vertices.

Figure 9(b) shows a set of (red) points finally generated by two (blue) bilaterally symmetrizable convex segments. Figure 10 shows two examples of applying bilateral symmetry: one for an on-ground range with a dominant segment, and the other for one with no dominant segment. The red-dotted try square-aligned segment in the on-ground range for the bed object in Fig. 10 is accepted, and two red reflected points are added to the set of generated points expected to be on the ground. No other segment is acceptable, and thus, the set of finally generated points consists of four convex vertices in the on-ground range and the two added points. On the one hand, there is no dominant segment in the on-ground range for the chair object in Fig. 10. The red-dotted try square-aligned (relatively long) segment is first accepted, and therefore, some red points are newly added to the set of generated points expected to be on the ground. Then the blue-dotted segment is also accepted, because all the points in the set fall inside the virtual shadow for the chair. We can see that the set of points generated finally is enough to estimate a meaningful base for the chair.

5.3 Estimation of a Base and Its Expansion

A set of the points expected to be on the ground has been ob-
tained from the test of symmetry. Then, a base can easily be estimated from the points, such that the base encloses them and has the same orientation as that of the try square-aligned segment. This alignment of orientation makes one of sides of the base collinear with a bilateral symmetry axis (that is, the bisector of the try square-aligned segment). Therefore, a bounding box with such a base looks natural. Figure 10 shows examples of such bases.

A new base is estimated whenever an on-ground range is given. Thus, there may be as many bases as the number of dominant segments, or the number of relatively long convex segments in the absence of dominant directions. Among these bases, the base with the minimum area is finally selected so that a 3D box bounding the target object as compactly as possible can be built up in later processing.

Sometimes, a finally selected base is too small to enclose the virtual shadow of an object when the object has a wider body than its ground section. Figure 9 (c) shows an example of such a case, where the blue rectangle represents the finally selected base and two blue-dotted lines are the support lines passing through the outermost two vertices of the base. The support lines lie between two red tangent support lines, and thus, pass through the virtual shadow. It is evident that the object corresponding to the virtual shadow cannot be packed into a bounding box with this base. Thus, this base needs to be expanded such that both the support lines passing through the outermost two vertices of the expanded base do not lie between two tangent support lines. Here, the base is scaled up, based on its center point, in the direction of the line connecting its outermost two vertices, as shown in Fig. 9 (c). Notice that the center of the base is preserved, even though one of the outermost vertices of the expanded base may exceed the range between two tangent support lines.

5.4 Determination of Height

A rectangular top for a bounding box with the base finally determined in the previous section has the same size and orientation as the base, and it is parallel to the ground plane. Thus, back-projection of the top onto the ground should be a rectangle with the same orientation as that of the base. Meanwhile, in the ground domain, each vertex of the rectangle should be on the support line of the corresponding vertex of the base, as shown in Fig. 9 (d), since the top is located vertically over the base.

On the one hand, the top of a bounding box for an image object needs to be determined, such that the shadow contour of the bounding box encloses the convex contour of the image object, as discussed in Property 4 in Sect. 3.2. Thus, the top contour in the shadow contour of the bounding box can be easily located in the ground domain by using the base contour determined in the previous section. The base contour consists of two perpendicular line segments meeting at a basis point. The top contour looks like the base contour rotated by 180°. The basis point of the top contour should lie on the support line of the rear right vertex of the base, as in Fig. 9 (d). When the top contour is properly located to minimally enclose the convex contour of the image object, the cross point of the top contour with the leftmost (or rightmost) support line can be utilized in computing the height of the bounding box. The top point \((x_t, y_t, h)\) corresponding to the cross point \((x_m, y_m, 0)\) should lie on the line passing through the camera origin \((x_c, y_c, z_c)\), and the cross point and its vertical projection onto the ground should be the base vertex \((x_b, y_b, 0)\) on the leftmost (or rightmost) support line (that is, \(x_t = x_b\) and \(y_t = y_b\)). Then, the height \(h\) can be computed as shown in Eqs. (4) and (5).

\[
(x_c - x_b, y_c - y_b, z_c - h) = k(x_c - x_m, y_c - y_m, z_c) \quad (4)
\]

\[
h = z_c(x_b - x_m)/(x_c - x_m) \text{ or } z_c(y_b - y_m)/(y_c - y_m) \quad (5)
\]

6. Experiment Results and Discussions

Two sets of objects are used in our experiment. A set of the 3D-modelled objects used to analyze the usefulness of the proposed observations was used again in order to measure the accuracy of the bounding boxes for various types of objects. Another set of 15 real objects found around us was used to test whether the proposed method, together with a calibrated camera, can determine appropriate bounding boxes for them, and to discuss the usefulness and limits of the proposed method in practical applications. All the image objects of the real objects are assumed to be extracted perfectly since the proposed method does not deal with segmentation of the image objects. Actually they were manually extracted in our experiment.

The resulting bounding boxes for 15 3D-modelled objects, nine real objects in a laboratory environment, two real outdoor objects, and two real objects in a testing center, are shown in Fig. 11, where each bounding box is superimposed on the corresponding image object. The bold lines in the displayed bounding boxes represent the dominant segments. The bounding boxes for the objects having dominant directions are well aligned to the dominant directions, and therefore look well fitted and natural. Usefulness of dominant
directions is also shown in Fig. 12. Two bounding boxes in the second column of Fig. 12, constructed by using the dominant directions shown in the first column, look very natural, while those in the last column, constructed using no information of the dominant directions, seem inadequate. On the one hand, for objects with no dominant direction in Fig. 11, each bounding box being aligned to a relatively long convex segment seems to be properly oriented and well fitted. For example, the barrel in Fig. 11 has no dominant direction and its round ground section causes the convex segments in its candidate convex sequence to be short. Nevertheless, some relatively long convex segments can be selected as the try square-aligned segments and then proper on-ground points can be determined after applying the observations and the bilateral symmetry. Thereby a base well fitted to the on-ground points can be estimated, as shown in Fig. 11.

Figure 13 shows the results of peak finding in the directional histograms for the ATM machine and the bed in Fig. 12. A peak in a directional histogram is a bin that satisfies two conditions: (1) it has the greatest value among its neighbors (±5 bins), and (2) its value is greater than a lower limit. The lower limit needs to be increased as the total number of edges, \( s \) (\( \Sigma \theta \)), in a directional histogram \( H(\theta) \) increases. On the one hand, we do not want to select dull peaks, which tend to occur more frequently as the standard deviation, \( \sigma \), of histogram values in \( H(\theta) \) decreases. Thus the lower limit \( t \) is designed to be proportional to the mean \( \mu \) of histogram values, but inversely proportional to the standard deviation, as shown in Eq. (6). The parameter \( k_0 \) was determined through experiments on a set of normalized directional histograms with \( s = 1 \). Notice that the lower limit \( t \) for a directional histogram \( H(\theta) \) with a total number of edges \( s \) is equal to \( s t_0 \) where \( t_0 \) is the lower limit of the normalized histogram \( H_0(\theta) \). Let the mean and the standard deviation for \( H_0(\theta) \) be \( \mu_0 \) and \( \sigma_0 \). Then \( t_0 \) is equal to \( \mu_0 + k_0/\sigma_0 \) since \( s = 1 \). On the one hand, the mean \( \mu \) and the standard deviation \( \sigma \) for \( H(\theta) \) can be written as \( s \mu_0 \) and \( s \sigma_0 \), respectively. Thus \( t = s t_0 \), as shown in Eq. (6).

\[
t = \mu + k \frac{s^2}{\sigma} = s \mu_0 + k_0 \frac{s^2}{s \sigma_0} = s \left( \mu_0 + k_0 \frac{1}{\sigma_0} \right) = s t_0 \quad (6)
\]

The red peaks in Fig. 13 correspond to dominant directions. The blue peak represents the vertical direction and thus cannot be a dominant direction. The black peak in the directional histogram for the bed represents another dominant direction, while the black peak in the one for the ATM machine should not be a dominant direction. However, we could not exclude the latter in our practical experiment, and thus, it was considered a dominant direction.

Figure 14 illustrates the usefulness of the bilateral symmetry property versus the rotational symmetry property. The left large image and the red rectangle are a virtual shadow of the chair in Fig. 10 and the base, as determined by using the bilateral symmetry property. A blue small base is also shown in Fig. 14, which is estimated from the same on-ground range by using the rotational symmetry property. The points expected to be on the ground are obtained by rotationally reflecting all the convex vertices in the on-ground range about the center point of the rightmost and leftmost green convex vertices. We can see that the base estimated by using the bilateral symmetry property is more adequate than the one from using the rotational symmetry property.

Accuracy of a bounding box for an object can be measured by comparing position, orientation, and size of the box with those of the object. The size of a bounding box can be described by its width, depth, and height. The position may be defined by the coordinates of the base center of the box, and the orientation by the angle between the x-axis of the world coordinate system and the line segment corresponding to width. In particular, the position, orientation, and size of a 3D-modelled object can be exactly computed from its modeling data. In our experiment, in order to obtain image objects with uniform size, 3D-modelled objects are normalized in size and then rendered from a virtual camera at a fixed position. The size of a 3D-modelled object is normal-
Table 3 Absolute mean errors in determining bounding boxes for 70 3D-modelled objects by using the proposed method and its partially modified methods; percentage (%) of object’s longest length (= 150) is used for the measurements having no unit.

| Methods                                      | Measurements |
|----------------------------------------------|--------------|
|                                             | Position (%) | Orientation (degree) | Width (%) | Depth (%) | Height (%) |
| The proposed method, P                       | 3.58         | 1.16                 | 6.90      | 6.33      | 4.00       |
| P without the concept of dominant directions | 4.13         | 1.70                 | 7.77      | 7.29      | 4.84       |
| P without applying bilateral symmetry        | 4.74         | 2.14                 | 8.21      | 7.53      | 5.64       |
| P with neither                               | 5.25         | 2.08                 | 9.28      | 8.15      | 5.49       |

Table 4 Accuracy of bounding boxes for nine different real objects in a laboratory environment.

|                | Absolute mean error | Standard deviation |
|----------------|---------------------|--------------------|
| Position (mm)  | 9.43                | 5.82               |
| Orientation (°) | 1.09                | 0.91               |
| Width (mm)     | 13.26               | 9.49               |
| Depth (mm)     | 17.86               | 8.45               |
| Height (mm)    | 18.69               | 14.03              |

Fig. 15 Comparison of accuracy among the proposed method and its three modified methods; colored circles represent absolute mean errors, and vertical line segments represent standard deviation.

ized such that the longest length among the width, depth, and height becomes 150. Table 3 shows the accuracy of the bounding boxes for the 70 3D-modelled objects. Notice that the units for position and size cannot be written because of the dimensionless definition of the 3D-modelled objects. We can see that almost all the bounding boxes are exactly oriented, because absolute mean error in orientation is very small. But accuracy of their size is not so good. The largest absolute mean error in width corresponds to 6.9% of the object’s longest length (= 150). Absolute mean errors in base area and volume correspond to 13.5% of the object base area and 13.8% of the object volume. On the other hand, absolute mean error in position corresponds to just 3.6% of the object’s longest length.

Table 3 and Fig. 15 show the usefulness of the concept of dominant directions and the bilateral symmetry property in the proposed method. Three partially modified methods were tested on a set of the 3D-modelled objects. They are 1) the proposed method without the concept of dominant directions, 2) one without applying bilateral symmetry, and 3) one with neither. We can see in Fig. 15 that the proposed method is the most accurate, and it outputs the most stable results. Therefore, we can conclude that the concept of dominant directions and the bilateral symmetry property are effective in determining bounding boxes.

The proposed method was also tested on nine different real objects in a laboratory, with a camera installed about 3 m high. Table 4 shows the accuracy of the bounding boxes for 19 image objects from the real objects. Average lengths of width, depth, and height for the real objects are 264.67 mm, 557.44 mm, and 450.11 mm, respectively. We can see that almost all the bounding boxes are also exactly oriented. They also seem to be exactly localized, because absolute mean error in position is just 9.43 mm. However, we have to notice that the distance between camera and real object is no more than 5 m. We can easily guess that the absolute mean error in position may increase as the distance increases.

Figure 16 shows another experiment with a real automated guided vehicle (AGV) that moves along a (blue) guideline drawn on the floor of a testing center. The camera was installed about 2.5 m high, and the AGV was moving at a distance of between 5 m and 12 m from the camera. We can see in Fig. 16 (a) that each bounding box of the AGV is properly determined, regardless of each pose of the AGV. The estimated position of the AGV from each bounding box in Fig. 16 (a) is marked with a yellow circle in Fig. 16 (b). Under the assumption that the AGV moves exactly along the guideline, accuracy of an estimated position is measured by using the distance between the estimated position and its nearest point on the guideline.

Table 5 shows the results of localization through 34 estimations shown in Fig. 16 (a). We can see that average absolute errors are not so bad when we consider the AGV's
Table 5 Accuracy of bounding boxes for a real AGV moving along a guideline on the ground.

|                          | Absolute mean error | Standard deviation |
|--------------------------|---------------------|--------------------|
| Position (mm)            | 39.94               | 29.30              |
| Width (mm)               | 113.60              | 41.94              |
| Depth (mm)               | 73.79               | 44.73              |
| Height (mm)              | 10.82               | 5.31               |

Fig. 17 Transitions of the positions estimated by three different methods: the proposed method, the center point-based method, and the lowest point-based method.

Another experiment was conducted on eight AGVs with various orientations at a fixed position in order to analyze how the orientation of an object affects estimation of position. The position of each AGV was estimated with three different methods: 1) the proposed method based on the base center point of the bounding box, 2) a method based on the lowest point of an image object, and 3) a method based on the center point of an image object. Figure 17 shows three transitions for the positions measured by those methods. Their standard deviations are 11.98 mm, 160.44 mm, and 16.35 mm, respectively. We can see that the transition by the proposed method is drastically stable, even though orientation of the AGV varies widely. From an average of eight positions determined by the proposed method, the position of the fixed point can easily be estimated. On the one hand, two transitions of the positions estimated by the specific point-based method, the center point-based method and the lowest point-based method, indicate that these methods are inadequate for positioning relatively large image objects in near- or mid-field scenes because they result in wrong positions and/or very large variations in position.

The proposed method was implemented under MATLAB code on a 1.86 GHz CPU. Table 6 shows the times needed for typical processes in the proposed method. Total average processing time for the 3D-modelled objects is greater than for the real objects. Thus, we can see that the 3D-modelled objects are finely designed and thus suitable for analyzing performance of the proposed method. On the one hand, we can see that standard deviation of processing time for the symmetry testing process is large. We found that the symmetry testing process for an image object with a round base required an extremely large amount of processing time, because there might be many bilaterally symmetrizable convex segments. Actually, in an experiment on the same data set without image objects with a round base, standard deviation of processing time for the symmetry testing process was reduced to 0.175 seconds, and thereby, average processing time was also reduced to 0.120 seconds. On the other hand, we also tested the processing times for the image objects of the nine real objects in the laboratory, which were automatically extracted by the background subtraction method [24]. We could find that the processing time for the automatically extracted image objects was almost similar to one for the manually extracted image objects.

Two inadequate bounding boxes and one abnormal bounding box are shown in Fig. 18. The inadequate bounding box for the object in Fig. 18 (a) results from incorrect estimation of on-ground points and improper application of bilateral symmetry, even though the object has a dominant direction. The bounding box for the 3D-modelled object of a computer monitor in Fig. 18 (b) seems to be too small to enclose the object, because the object is not very good for guessing 3D volume. On the one hand, the image object of the AGV in Fig. 18 (c) represents a degenerated case. A 2D

Table 6 Analysis of processing time in each process of the proposed method (unit: seconds).

| Processes                        | Objects  | 3D-modelled objects (average) | Real objects (average) | All objects | Avg. | Std. Dev. |
|----------------------------------|----------|-------------------------------|------------------------|-------------|------|-----------|
| Domain transformation            |          | 0.279                         | 0.256                  | 0.272       | 0.064|
| Determining dominant directions  |          | 0.272                         | 0.248                  | 0.265       | 0.040|
| Finding on-ground range          |          | 0.007                         | 0.005                  | 0.007       | 0.005|
| Testing symmetry                 |          | 0.213                         | 0.134                  | 0.190       | 0.457|
| Determining a bounding box       |          | 0.009                         | 0.008                  | 0.009       | 0.001|
| Total average time               |          | 0.780                         | 0.651                  | 0.743       | -    |

Fig. 18 Examples of inadequate or abnormal cases: (a) a view inadequate to estimate on-ground points, (b) an object where 3D volume is hard to guess, and (c) a degenerated view.
base shape could not be generated because the on-ground range estimated from the image object looked just like a line segment.

7. Conclusions

A novel method for determining a bounding box to estimate 3D pose (position and orientation) and 3D size for a segmented object from a single image is proposed and evaluated in this paper. The method is designed to work on an object lying on a known plane through camera calibration, such as ground under surveillance or a work floor, and to determine a bounding box with orientation aligned to one of the dominant directions of the object. The constraint on the objects on a known plane allows the number of degrees of freedom in localizing a bounding box in 3D world space to be reduced by 3, while the alignment to a dominant direction by 1.

The method first estimates the ground portion of a given image object by using some observations on structural characteristics of general objects, and then a suitable base is determined by using an assumption of bilateral symmetry in objects that is useful for recovering the self-occluded ground portion. A bounding box with the base is finally constructed by determining its height, such that back-projection of the constructed box onto the ground minimally encloses that of the given image object.

Performance of the method was analyzed through experiments with 3D-modelled objects and real objects. Bounding boxes aligned to one of the dominant directions look natural and well-fitted. We could see, from experiments with 3D-modelled objects, that the linearly ordered observations on structural characteristics of objects are very useful in determining the ground portion of a given image object, even though there were a few unavoidable misinterpretations. Recovery of the self-occluded backside of the ground portion based on bilateral symmetry of dominant segments and/or relatively long segments was also useful. However, in experiments with the 3D-modelled objects, absolute mean error in base area corresponding to about 13.5% of the object’s base area was mostly caused from this recovery, with an error of about 15.9% from recovery based on rotational symmetry. Fortunately, absolute mean error in position was not very large, so the proposed method may be used for positioning target objects with no marker. Actually, in the experiment with real objects and a real AGV, the proposed method localized them with a high degree of accuracy. The height of bounding boxes for box-like objects was also determined relatively well, but the height for objects with a narrow or unbalanced upper part, such as a cone or a chair, was estimated to be lower than the real height.

The proposed method is expected to be used effectively in the fields of autonomous navigation and smart surveillance. We hope that future research can improve the proposed observations and the use of symmetry through novel points of view on features of objects.

Acknowledgments

This work was supported by a 2-Year Research Grant of Pusan National University.

References

[1] M.H. Sedky, M. Moniri, and C.C. Chibelushi, “Classification of smart video surveillance systems for commercial applications,” Proc. IEEE Conf. on Advanced Video and Signal based Surveillance, pp.638–643, Sept. 2005.
[2] R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wilson, “A system for video surveillance and monitoring,” VSAM Final Report, CMU-RI-TR-00-12, The Robotics Institute, Carnegie Mellon University, 2000.
[3] W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE Trans. Syst. Man Cybern. C: Applications and Reviews, vol.34, no.3, pp.334–352, Aug. 2004.
[4] T.N. Tan, G.D. Sullivan, and K.D. Baker, “Model-based localisation and recognition of road vehicles,” Int. J. Comput. Vis., vol.27, no.1, pp.5–25, March 1998.
[5] Z. Li, J. Liu, and X. Tang, “A closed-form solution to 3D reconstruction of piecewise planar objects from single images,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol.1, pp.17–22, June 2007.
[6] W. Ruisheng and F.P. Ferrie, “Camera localization and building reconstruction from single monocular images,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshop, pp.23–28, June 2008.
[7] S. Jung, D. Jang, and M. Kim, “3D localization of projected objects for surveillance,” Proc. Third ACM/IEEE Conf. on Distributed Smart Cameras, CD Proceeding (ISBN:978-1-4244-4620-9), Aug. 2009.
[8] G. Barrequet and S. Har-Peled, “Efficiently approximating the minimum-volume bounding box of a point set in three dimensions,” J. Algorithms, vol.38, no.1, pp.91–109, Jan. 2001.
[9] M. Kim, C. Park, and K. Koo, “Natural/man-made object classification based on Gabor characteristics ,” Image and Video Retrieval, Lect. Notes Comput. Sci., vol.3568, pp.550–559, July 2005.
[10] M. Kim, C. Park, and K. Koo, “On object classification: Artificial vs. natural,” Image Analysis and Recognition, Lect. Notes Comput. Sci., vol.3656, pp.532–539, Sept. 2005.
[11] J. Rosen, Symmetry discovered: Concepts and applications in nature and science, Courier Dover Publications, 1975.
[12] S. Thrun and B. Wegbreit, “Shape from symmetry,” Proc. Tenth IEEE International Conf. on Computer Vision, vol.2, pp.1824–1831, Oct. 2005.
[13] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.11, pp.1330–1334, Nov. 2000.
[14] H. Zabrodsky and D. Weinshall, “Using bilateral symmetry to improve 3D reconstruction from image sequences,” Comput. Vis. Image Understand., vol.67, no.1, pp.48–57, July 1997.
[15] R. Hartley and A. Zisserman, Multiple view geometry in computer vision 2nd ed., Cambridge University Press, New York, 2003.
[16] P.V.C. Hough, Methods and means for recognizing complex patterns, U.S. Patent 3,069,654, filed March 1962, issued Dec. 1962.
[17] R.C. Gonzalez and R.E. Woods, Digital image processing 2nd ed., Prentice Hall, New Jersey, 2002.
[18] H. Freeman and R. Shapira, “Determining the minimum-area enclosing rectangle for an arbitrary closed curve,” Commun. ACM, vol.18, no.7, pp.409–413, July 1975.
Sunghoon Jung 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.

Minhwan Kim 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.

[19] G.T. Toussaint, “Solving geometric problems with the rotating calipers,” Proc. IEEE Mediterranean Electrotechnical Conference ’83, pp.24–26, May 1983.
[20] 3D-modelled objects, http://archive3d.net, accessed Nov. 2013.
[21] R.P. Grimaldi, Discrete and combinatorial mathematics 5th ed., Pearson, 2006.
[22] C.B. Barber, D.P. Dobkin, and H.T. Huhdanpaa, “The quickhull algorithm for convex hulls,” ACM Trans. Mathematical Software, vol.22, no.4, pp.469–483, Dec. 1996.
[23] H. Koyuncu and S.H. Yang, “A survey of indoor positioning and object locating systems,” Int. J. Computer Science and Network Security, vol.10, no.5, pp.121–128, May 2010.
[24] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts, and shadows in video streams,” IEEE Trans. Pattern Anal. Mach. Intell., vol.25, no.10, pp.1337–1342, Oct. 2003.