December 1973

OBJECT RECOGNITION IN MULTISENSORY SCENE ANALYSIS

by

David Nitzan

Artificial Intelligence Center

Technical Note 83

SRI Project 1530

The research reported herein was supported at SRI by the Advanced Research Projects Agency of the Department of Defense, monitored by the U.S. Army Research Office, Durham, North Carolina, under Contract DAHC04-72-C-0008.

ABSTRACT

A perception strategy for recognizing an object on the basis of color and range data is illustrated by finding a desk in an office scene. Geometrical models of objects (such as a desk, a chair, and a glass) are described by transformation (using scaling, rotation, and translation matrices) of primitive surfaces (unit cube, sphere, and cylinder). An image-bounding window for the acquisition element of an object is based on range data. Image bounding windows are derived for four cases in which single and multiple acquisition elements are oriented vertically and horizontally.

CONTENTS

1 STRATEGY

The work of the SRI vision group on a knowledge-base perception system was described by Tenenbaum.[1]* This note describes geometrical models of objects that may be found in indoor scenes. By way of introduction, the strategy for recognizing an object on the basis of color and range data is described first, using bottom-up and top-down procedures.

A general hierarchical perception model, based on the color and geometry of object surfaces, is shown in Figure 1. Given the parameters of such a model, one of the goals of the perception system is to find a specified object in the scene. The strategy to achieve this goal is dynamic in the sense that alternative subroutines may be called by a higher level program. These include the COLOR subroutine, which determines the color of a given surface element from color-filtered intensity data, and the following geometry subroutines, which are based on range data:

(1) FLOOR subroutine, which finds all the image points belonging to the floor.[2]

(2) HORIZONTAL PLANE subroutine, which finds the sets of all the image points belonging to horizontal planes of given heights and tolerances.[2]

(3) SURFACE BOUNDARY subroutine, which finds all the image points surrounding individual bodies.[2]

(4) INTERSECTION BOUNDARY subroutine, which finds the image points of intersection boundaries between surface boundaries.[2]

---

*Superscripts denote references listed at the end of this Technical Note.

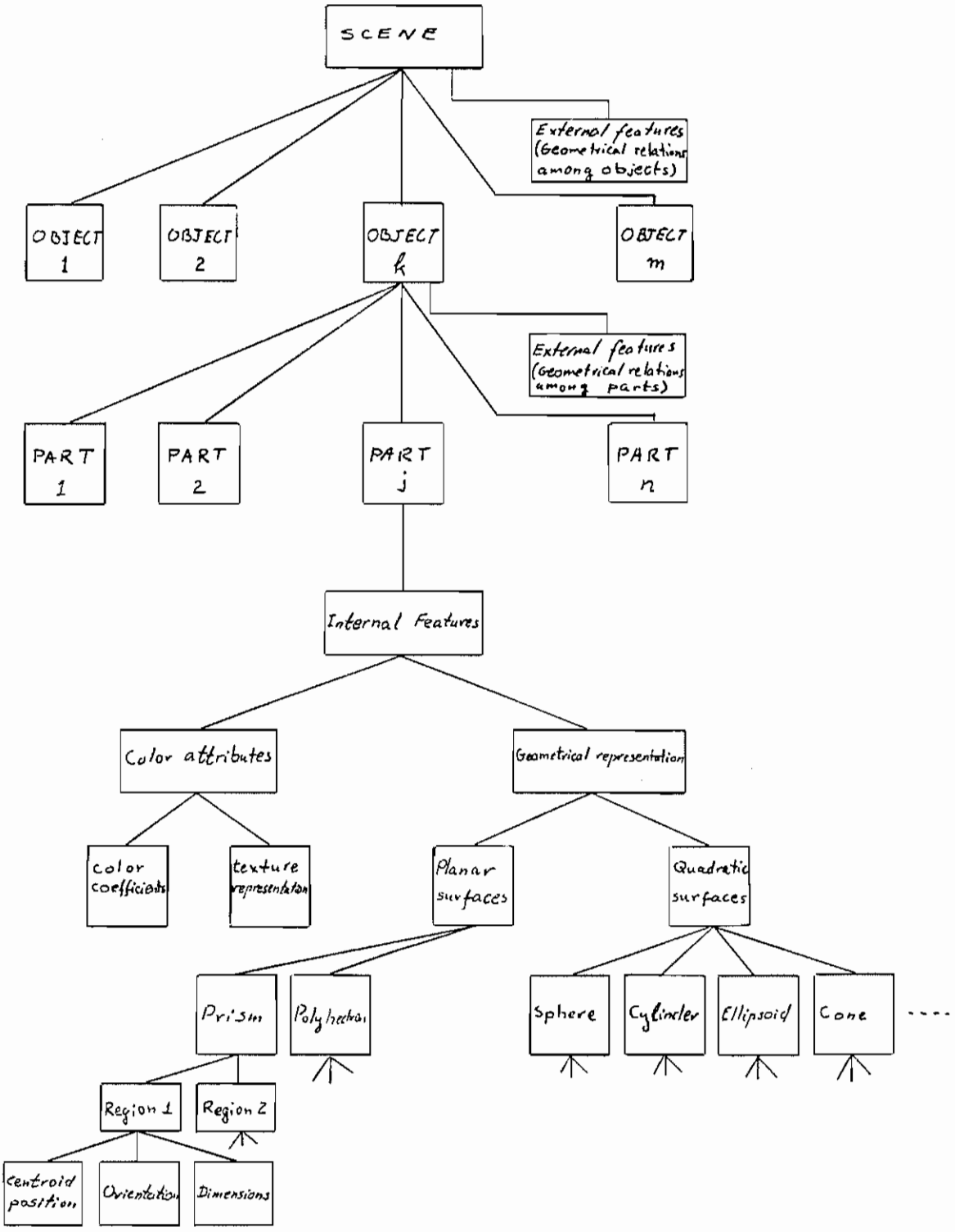


FIGURE 1 GENERAL HIERARCHICAL PERCEPTION MODEL

(5) ORIENTATION subroutine, which computes the orientation of a given surface element.

(6) BOUNDING WINDOW subroutine, which determines an image polygon where the image of an object, an object part, or a part region might be found.

(7) PLANE FIT subroutine, which finds the image points of a planar region to which a given surface element belongs.

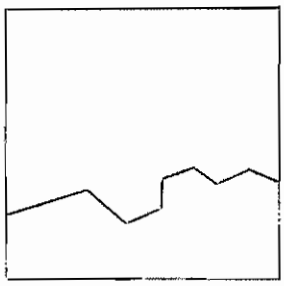
Both top-down and bottom-up procedures will be employed, depending on their figure of merit. For example, the FLOOR subroutine, basically a bottom-up procedure, may be called first as part of a top-down procedure to find an object near the floor, such as baseboard, waste basket, chair, or desk. Similarly, the HORIZONTAL PLANE subroutine may be called as an initial step for finding a table, a desk, or a chair.

The basic top-down procedure consists of sampling elements and comparing their geometrical and color attributes with those of the object to be found. It is executed along the steps described below in the example of finding a desk and is illustrated in Figure 2.

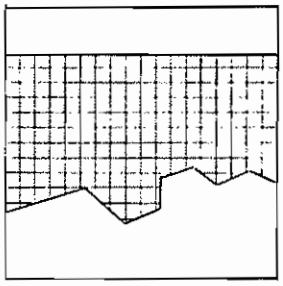
1.1 Step 1--Acquisition Bounding Window

The search for an acquisition element that matches given criteria may be reduced if there is sufficient information for roughly bounding the window where such an element may be found. For example, once the floor image, Figure 2(a), has been found, the window for an acquisition element belonging to a desk is the largest region above the floor that can accommodate the highest point of a desk top at the farthest point in the scene (e.g., near the wall), as illustrated in Figure 2(b).* Alternatively, the bounding window may be fixed by the floor boundary and

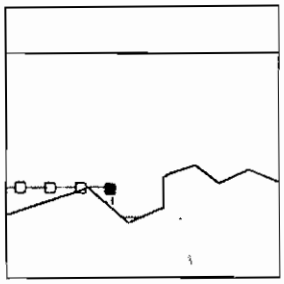
*The same bounding window will be used, if necessary, to limit the search for the horizontal desk top (by calling the HORIZONTAL PLANE subroutine) or the surface-boundary points associated with the desk (by calling the SURFACE BOUNDARY subroutine).



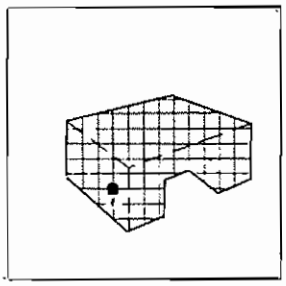
(a) EXTRACTING THE FLOOR REGION



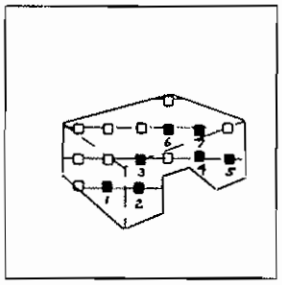
(b) ACQUISITION BOUNDING WINDDW



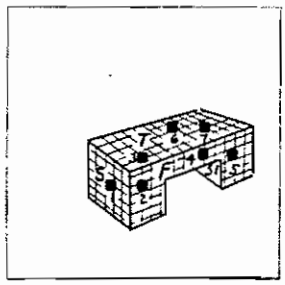
(c) ACQUISITION ELEMENT SEARCH BY COARSE SAMPLING



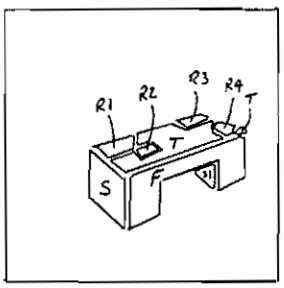
(d) BOUNDING WINDOW FOR DESK



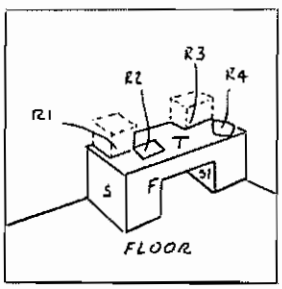
(e) COMPLETED COARSE SAMPLING



(f) MODIFIED BOUNDING WINDOW FOR DESK SUBDIVIDED INTO TOP, SIDE, AND FRONT REGIONS



(g) RECOGNITION OF TOP, SIDE, AND FRONT REGIONS. REGIONS R1-R4 DO NOT BELONG TO DESK.



(h) FINAL INTERPRETATION OF DESK

FIGURE 2 FINDING A DESK BY SAMPLING PROCEDURE

Solid squares indicate surface elements that match orientation and color attributes of a desk model.

the picture frame above the floor. If the information for limiting the search is not available, then the window extends to the entire N × N picture points.

1.2 Step 2--Searching an Acquisition Element

A coarse sampling scan is performed within the acquisition bounding window in the search for a surface element (e.g., a 3 × 3 square) whose color and orientation attributes match those of the object to be found. Depending on the situation, the matching attributes may characterize the surface of one of the following:

- A region of a part of the object
- A part of the object (i.e., any region of the part)
- The object (i.e., any region of any part of the object).

The sampling interval ΔJ (or ΔI) of the horizontal (or vertical) scan will correspond to the image of one-half or one-third of the smallest size of the largest surface (of the region, part, or object under search) at the farthest range within the assumed environment. Since the image of an object increases as the magnitude of the tilt angle relative to the horizon increases, the search scan for regions below the horizon will proceed upward (decreasing I values), and vice versa. In the example shown in Figure 2(c), an acquisition element belonging to any region of the desk is searched upward. A vertical element of the proper color (solid square Number 1) is found and is associated with the desk side.

1.3 Step 3--Surface Bounding Window

Having found an acquisition element, we now determine a second, smaller bounding window for the image of the bounding space where the surface of the desired object might possibly be located [see Figure 2(d)]. The bounding space depends on the internal and external geometrical

attributes of the object, the vertical direction of the acquisition scan (upward or downward), the orientation of the acquisition element, and any surface-boundary points obtained by calling the SURFACE BOUNDARY subroutine. The effects of these factors on the bounding space are discussed in Section 4.

Next, the coarse sampling scan of Step 2 is continued within the surface bounding window, as illustrated in Figure 2(e). This is done for two reasons:

- To verify the first acquisition element. (If no additional elements are acquired, then the first acquisition element is assumed to be erroneous due to noise.)
- On the basis of the additional information, to reduce the surface window further and subdivide it into subwindows for large regions. The subwindows may overlap by different amounts. In Figure 2(f), for example, there is practically no overlap. Clearly, the smaller the overlap, the better.

1.4 Step 4--Surface Recognition

Recognition of the remainder of the surface to be found is now performed. As in Step 2, three cases are distinguished with regard to recognition of either a region or a part or the entire object. The recognition process is similar to the sampling search in Step 2, except for the following:

(1) The sampling interval is smaller. In most cases it will be equal to the element size, i.e., the scanning will cover the entire area of the modified (smaller) surface bounding window obtained by the end of Step 3.

(2) The attributes to be compared and the order of comparison will depend on the region subwindow, provided there is little overlap among the subwindows, as is the case in Figure 2(f).

Suppose first that the overlap among the region subwindows is large and hence the entire object is to be recognized. In such a case, the attributes of all the object regions are compared simultaneously with the attributes of each element. The element is then classified into one of four categories:

(1) It does not belong to the object.

(2) It belongs to one region of the object.

(3) It belongs to one of at least two regions of the object (or another object).

(4) It belongs to at least two regions, some of which may be part of another object.

If the attributes of an element are not homogeneous, it is classified into Category (4), no comparison is made, and the next element is examined. Otherwise, the attributes are compared and the element is classified into Category (1), (2), or (3) if there is a match with none, one, or more than one region, respectively. The elements under Categories (1) and (2) are merged into the corresponding individual regions. The resulting regions may help the labeling and merging of some of the elements under Category (3) on the basis of the model region attributes (shape, size, and the like). The remaining, unlabeled elements under Category (3) and the elements under Category (4) may next be subdivided into finer elements and the above procedure is repeated, if more exact region boundaries are required.

The above procedure is also applicable to finding a part of an object. This case may be viewed as finding a one-part object. Furthermore, the same procedure is also applicable to finding a region, except that it is simpler because:

(1) Category (3) does not exist

(2) The number of attributes to be compared with is smaller.

In the example illustrated in Figure 2(g), the desk-top region T, the side regions S and S1, and the front region F have been found by merging all the elements under Category (2), and Regions R1 through R4 have been classified under Category (1). The unmarked strips between these regions consist of elements belonging to Categories (3) and (4). These strips are shrunk by using smaller elements and, using information about the shape and size of a desk, the final interpretation shown in Figure 2(h) is reached.

2 OBJECT MODEL

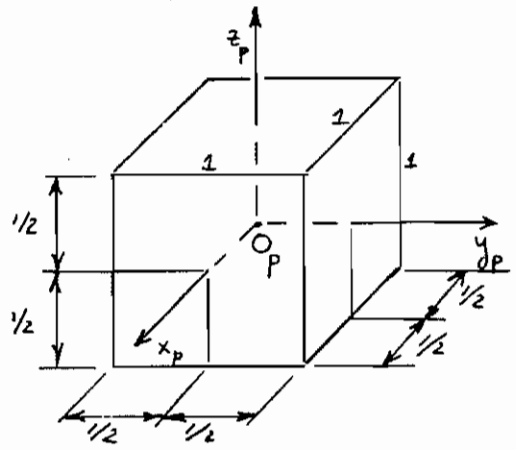
A right-hand Cartesian coordinate system (x_m, y_m, z_m) will be assumed to be the reference frame for describing the geometrical attributes of each object model, such as TABLE, DESK, CHAIR, BASKET, ARM, GLASS, and so on. In Figure 1 the model consists of parts, each of which is described by color attributes and geometrical representation. The latter consists of planar surfaces and quadratic surfaces. Planar surfaces are classified into prisms in particular and into polyhedra in general. Quadratic surfaces are classified into spheres, cylinders, ellipsoids, cones, and so on.

2.1 Primitive Surfaces

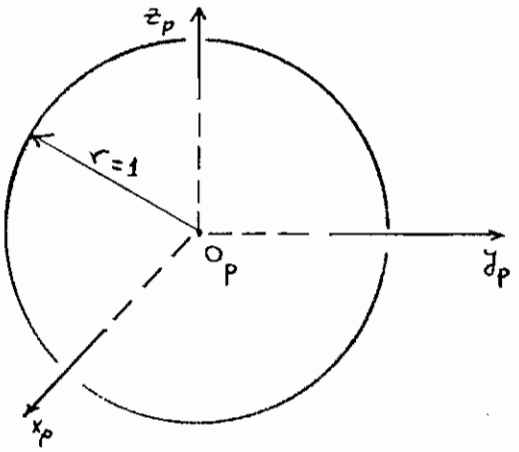
Each surface will be described by scaling, rotating, and translating a primitive surface within the (x_m, y_m, z_m) reference frame. Three primitive surfaces, a unit cube, a unit sphere, and a unit cylinder, are shown in Figure 3. Note that an ellipsoid may be described by scaling a sphere by different amounts in different directions.

Other primitives may be added, such as a wedge and a cone. Further extension of the primitives is possible by allowing half of them to be referred to, e.g., a hemisphere or half a cylinder (of D-shape cross section).

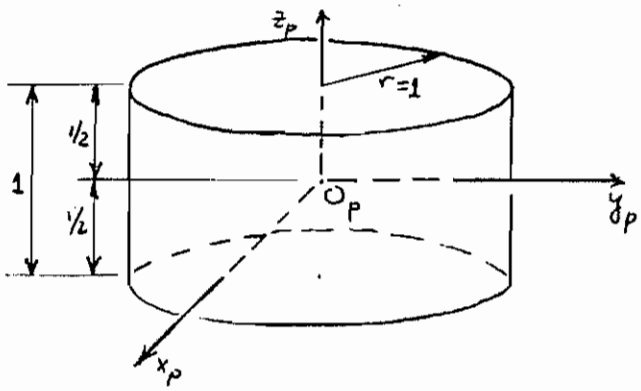
An object surface will be designated according to its primitive, i.e., PRj (Prism j), SPj (Sphere j), CYj (Cylinder j), HSj (Hemisphere j), COj (Cone j), ELj (Ellipsoid j), and so on.



(a) A UNIT CUBE



(b) A UNIT SPHERE



(c) A UNIT CYLINDER

FIGURE 3 PRIMITIVE SURFACES

2.2 Transformation Matrices

To model an object surface, a primitive surface is first scaled, then rotated about the primitive origin O_p, and finally translated in the model coordinates (x_m, y_m, z_m). Using homogeneous coordinates, scaling is performed by the matrix S, rotation by the matrix R, and translation by the matrix T, as follows:

2.2.1 Scaling

The $(x_p, y_p, z_p)^t$ coordinates of any point on a primitive surface are first scaled into $(x_s, y_s, z_s)^t$ coordinates by scale factors s_x, s_y, s_z, respectively:

$$\tilde{v}_s = \begin{bmatrix} wx_s \\ wy_s \\ wz_s \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} wx_p \\ wy_p \\ wz_p \\ w \end{bmatrix}}_{\tilde{v}_p} = S\,\tilde{v}_p \quad . \qquad (1)$$

2.2.2 Rotation

The $(x_s, y_s, z_s)^t$ coordinates of any point on the scaled surface are next rotated into $(x_r, y_r, z_r)^t$ coordinates by angles α_x, α_y, and α_z around each axis, respectively:

$$\tilde{v}_r = \begin{bmatrix} wx_r \\ wy_r \\ wz_r \\ w \end{bmatrix} = R_\ell\, R_k\, R_j \underbrace{\begin{bmatrix} wx_s \\ wy_s \\ wz_s \\ w \end{bmatrix}}_{\tilde{v}_s} = R_{jk\ell}\,\tilde{v}_s \quad , \qquad (2a)$$

where jkℓ is the order of rotation, j, k, or ℓ standing for x, y, or z; and where

$$\left.\begin{aligned} R_x &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha_x & \sin\alpha_x & 0 \\ 0 & -\sin\alpha_x & \cos\alpha_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ R_y &= \begin{bmatrix} \cos\alpha_y & 0 & -\sin\alpha_y & 0 \\ 0 & 1 & 0 & 0 \\ \sin\alpha_y & 0 & \cos\alpha_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ R_z &= \begin{bmatrix} \cos\alpha_z & \sin\alpha_z & 0 & 0 \\ -\sin\alpha_z & \cos\alpha_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}\right\} . \qquad (2b)$$

2.2.3 Translation

The $(x_r, y_r, z_r)^t$ coordinates of any point on the scaled and rotated surface are finally translated into the model coordinates $(x_m, y_m, z_m)^t$ by the amounts x_c, y_c, z_c:

$$\tilde{v}_m = \begin{bmatrix} wx_m \\ wy_m \\ wz_m \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & x_c \\ 0 & 1 & 0 & y_c \\ 0 & 0 & 1 & z_c \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} wx_r \\ wy_r \\ wz_r \\ w \end{bmatrix}}_{\tilde{v}_r} = T\,\tilde{v}_r \quad . \qquad (3)$$

Combining Eqs. (1) through (3), we obtain

$$\tilde{v}_m = \begin{bmatrix} wx_m \\ wy_m \\ wz_m \\ w \end{bmatrix} = T\, R_{jk\ell}\, S \begin{bmatrix} wx_p \\ wy_p \\ wz_p \\ w \end{bmatrix} , \qquad (4)$$

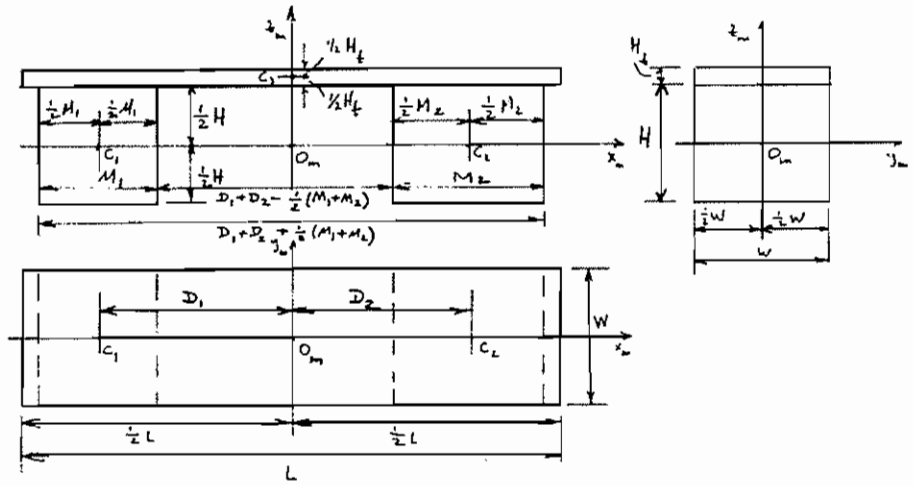
where $R_{jk\ell} = R_\ell\, R_k\, R_j$. The nine input parameters associated with each surface are given in the parameter array

$$P = \begin{pmatrix} s_x & s_y & s_z \\ \alpha_x & \alpha_y & \alpha_z \\ x_c & y_c & z_c \end{pmatrix} . \qquad (5)$$

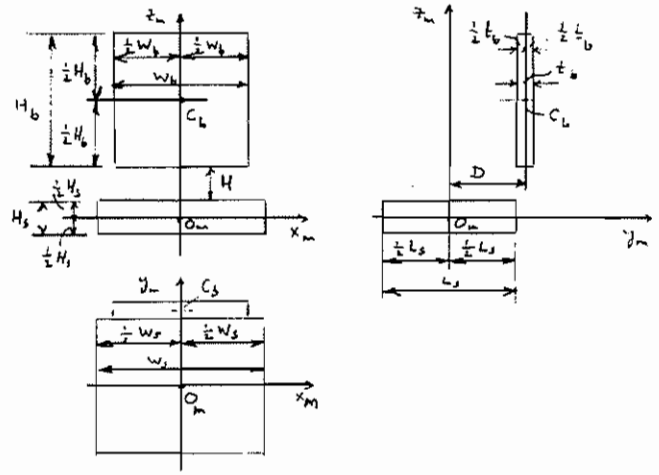
An object part or a one-part object may consist of more than one surface. The partition of an object into parts is implicit by the values in the parameter arrays of the object surfaces. Additional rotation and translation transformations will be provided to describe any geometrical constraints among objects. The parameters of these transformations can be derived from the geometrical models (see Figure 4) and statements describing such constraints as direct support (e.g., "GLASS ON DESK" or "DESK ON FLOOR"), indirect support (e.g., "CHAIR SEAT 18 INCHES ABOVE FLOOR"), orientation (e.g., "DESK PARALLEL WALL"), and so on.

2.3 Examples

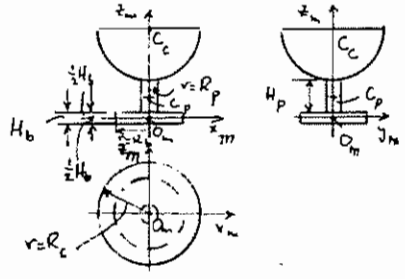
Three geometrical models are illustrated in Figure 4 for a desk, a chair (seat and back only), and a (champagne) glass. The surface names (corresponding to their primitives) and the parameter arrays associated with each of these models are given as follows.



(a) DESK



(b) CHAIR



(c) GLASS

FIGURE 4 GEOMETRICAL MODELS

2.3.1 Desk

$$P_{desk,1} \equiv PR1 = \begin{pmatrix} M_1 & W & H \\ 0 & 0 & 0 \\ -D_1 & 0 & 0 \end{pmatrix} . \quad (6)$$

$$P_{desk,2} \equiv PR2 = \begin{pmatrix} M_2 & W & H \\ 0 & 0 & 0 \\ D_2 & 0 & 0 \end{pmatrix} . \quad (7)$$

$$P_{desk,top} \equiv PR3 = \begin{pmatrix} L & W & H_t \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\left(H + H_t\right) \end{pmatrix} . \quad (8)$$

2.3.2. Chair

$$P_{chair,back} \equiv PR1 = \begin{pmatrix} W_s & L_s & H_s \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \quad (9)$$

$$P_{chair,back} \equiv PR2 = \begin{pmatrix} W_b & t_b & H_b \\ 0 & 0 & 0 \\ 0 & D & H + \frac{1}{2}\left(H_s + H_b\right) \end{pmatrix} . \quad (10)$$

2.3.3. Glass

The glass model consists of two cylinders whose parameter arrays are

$$P_{glass,base} \equiv CY1 = \begin{pmatrix} R_b & R_b & H_b \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad (11)$$

$$P_{glass,pole} \equiv CY2 = \begin{pmatrix} R_p & R_p & H_p \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\left(H_b + H_p\right) \end{pmatrix} , \qquad (12)$$

and a hemisphere whose parameter array is

$$P_{glass,cup} \equiv HS1 = \begin{pmatrix} R_c & R_c & R_c \\ 0 & 0 & 0 \\ 0 & 0 & R_c + H_p + \frac{1}{2} H_b \end{pmatrix} . \qquad (13)$$

2.4 Modeling of Similar Objects

Suppose that there are two identical objects in the scene, OBJ1 and OBJ2. For convenience of a user (programmer), there is no need to enter the object description twice. Instead, a provision will be made to enter the description "OBJ2 IS IDENTICAL TO OBJ1" or, in short, "OBJ2 IDENTICAL OBJ1" or "OBJ2 = OBJ1."

In some cases two objects in the scene will be similar but not identical. Here, again, there is no need for the entire model description. Instead, the name of the similar object and values of different parameters will be entered. For example, consider two desks: DESK1, whose parameter arrays are given in Eqs. (6) through (8), and DESK2, which is similar to DESK1 except that $M_1 = M_2 \equiv M$ and $D_1 = D_2 \equiv D$. The description of DESK2 may then be "DESK2 SIMILAR DESK1 EXCEPT PR1($s_x = M, x_c = -D$), PR2($s_x = M, x_c = D$)." This method may be used even if the objects are not "similar" to the human eye, but their models use common primitives. For example, DESK2 may be used to describe TABLE1 by specifying different parameter values. Similarly, the GLASS model in Figure 4(c) and Eqs. (11) through (13) may be modified to describe a LAMP model.

3 ACQUISITION BOUNDING WINDOW

Bottom-up information, such as the floor region, may limit the search for an acquisition element. Consider, for example, the acquisition of a desk and assume that the floor region has been acquired at the bottom of the picture with no "holes" in it, as shown in Figure 5(a). Let Point F, where $I = I_F$ and $J = J_F$, designate the "highest" floor point, i.e., whose I value is minimal. The acquisition bounding window is bounded by the floor at the bottom and the row $I = I_M$ on top. We wish to find the minimum value of I_M, i.e., the safest acquisition bounding window.

By modifying Eq. (18) in Chapter 10 of Duda and Hart,[3] the image coordinates (x_p, z_p) of a point P(x, y, z) in space are as follows:

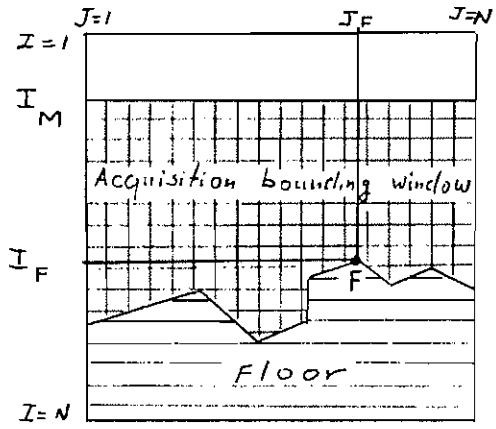
$$x_p = f_r \frac{x}{y \cos \varphi_0 + \left(z_0 - z\right) \sin\left(-\varphi_0\right)} \quad , \tag{14}$$

$$z_p = f_r \frac{y \sin\left(-\varphi_0\right) - \left(z_0 - z\right) \cos \varphi_0}{y \cos \varphi_0 + \left(z_0 - z\right) \sin\left(-\varphi_0\right)} \quad , \tag{15}$$

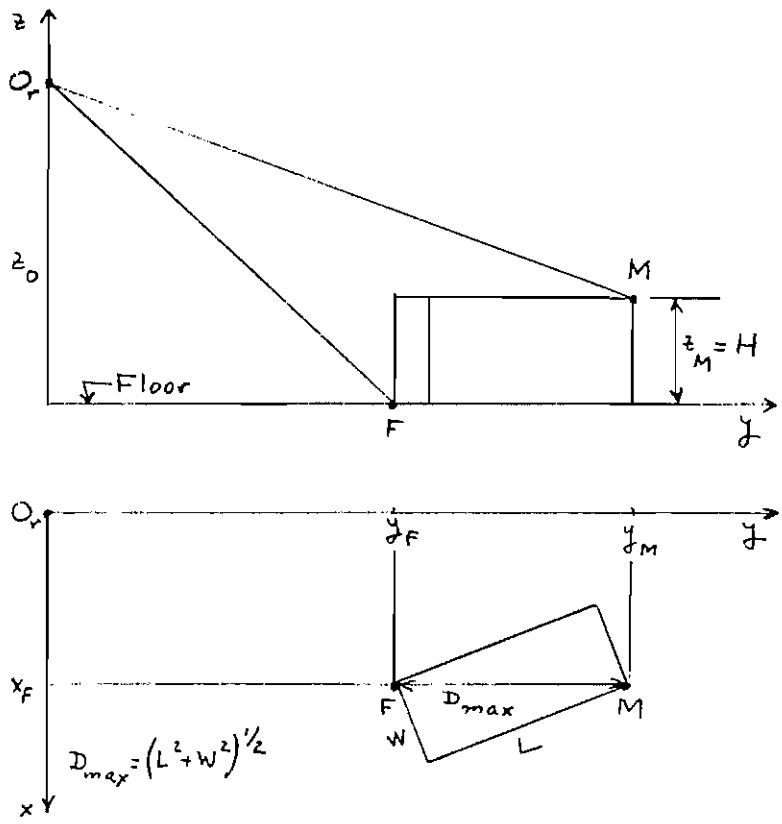
where φ_0 is the tilt angle of the camera and f_r is the distance from the lens center to the image plane. Rearranging Eq. (15),

$$z_p = f_r \frac{u \tan\left(-\varphi_0\right) - 1}{u + \tan\left(-\varphi_0\right)} \quad , \tag{16}$$

where $u = y/(z_0 - z)$. The function $z_p(u)$ is hyperbolic with no maximum: the higher u, the higher z_p. The value of $I = I_M$, therefore, corresponds to a desk point $M(x_M, y_M, z_M)$ whose $y/(z_0 - z)$ is maximal, i.e., whose z and y values are both maximal. In the desk model, Figure 4(a), the



(a) FLOOR REGION AND ACQUISITION BOUNDING WINDOW



(b) FRONT AND TOP VIEW OF DESK MODEL IN WORST POSITION

FIGURE 5 DETERMINATION OF ACQUISITION BOUNDING WINDOW FROM FLOOR REGION AND DESK MODEL

maximal value of z is $z_M = H$. In Figure 5(b), the maximal value of y is obtained by positioning the desk so that its diagonal top view dimension, $D_{max} = (L^2 + W^2)^{\frac{1}{2}}$, is colinear with the y axis, overlapping Point F at one end. Denoting the global coordinates of Point F by $(x_F, y_F, 0)^t$, then

$$\left.\begin{aligned} x_M &= x_F \\ y_M &= y_F + D_{max} \\ z_M &= H \end{aligned}\right\}. \qquad (17)$$

Substituting $y = y_M$ and $z = z_M$ of Eq. (17) into Eq. (15), we obtain

$$z_{pM} = f_r \frac{\left(y_F + D_{max}\right) \tan\left(-\varphi_0\right) - \left(z_0 - H\right)}{y_F + D_{max} + \left(z_0 - H\right) \tan\left(-\varphi_0\right)} \quad . \qquad (18)$$

The corresponding row index I_M is obtained by substituting $z_p = z_{pM}$ into the relation I versus z_p. [See Eq. (1.8) in Ref. 2.] Thus,

$$I_M = \frac{N + 1}{2} - \frac{N - 1}{2\, z_{pm}} z_{pM} \quad , \qquad (19)$$

where $-z_{pm} \leq z_p \leq z_{pm}$. Obviously, if $I_M \leq 1$, then $I_M = 1$.

So far we have assumed that the floor region covers the bottom of the view. This, of course, is not always the case. A desk top may be seen in the foreground of the view, "below" the floor. Thus, any non-floor region whose $I > I_M$ is part of the acquisition bounding window.

In conclusion, to determine the acquisition bounding window:

(1) Find the floor image points and determine I_F and J_F for which I is minimal.

(2) Using the range analysis in Ref. 2, compute x_{pF} and z_{pF} from Eqs. (1.7) and (1.8), and then compute y_F from Eq. (2.2).*

(3) Substituting y_F and the maximal (diagonal) dimension, D_{max}, in the model of the object to be acquired, compute z_{pM} and I_M from Eqs. (18) and (19).

(4) The acquisition bounding window is obtained by deleting the floor points and the points for which $1 \leq I < I_M$ from the entire N × N array.

---

*In Ref. 2, see p. 9 for Eqs. (1.7) and (1.8) and p. 16 for Eq. (2.2).

4 SURFACE BOUNDING WINDOW

Determination of the surface bounding window is described in this section. First, we define the bounding space, discuss the factors affecting this space, propose means to represent it, and show how to match it to an acquisition element and determine the (image) bounding window. Second, we deal with a bounding space corresponding to a single acquisition element and distinguish between a vertical and a horizontal element. Finally, the above topic is discussed for multiple acquisition elements.

4.1 Bounding Space

A bounding space is the three-dimensional space where an object model might be located, given an acquisition sample (or more than one sample) of the object. The bounding space depends on four factors:

(1) The geometrical model of the object.

(2) The external geometrical relations (constraints) among objects in the scene.

(3) The information carried by the acquisition element(s).

(4) The direction of the acquisition scanning (decreasing or increasing I values).

A bounding space may be described like an object model, i.e., by means of planar and quadratic primitive surfaces and the transformation matrices S (Scaling), R (Rotation), and T (Translation), using homogenous coordinates. The centroid of the acquisition element will be the origin O_b of the reference frame for the bounding space. The reference frame will be a right-hand Cartesian coordinate system, to be designated by (x_b, y_b, z_b).

Each bounding space of an object (or an object region) will be represented by a set of extreme points that envelop the bounding space. Different sets of n_x extreme points will be used, depending on each of the four factors listed above. Several alternative ways are considered for describing the extreme points:

(1) Listing the coordinates (x_{bj}, y_{bj}, z_{bj}) of every extreme point, where $j = 1, 2, \ldots, n_x$.

(2) Same as Item (1), except that the coordinates will be computed, using a given set of expressions.

(3) Same as Item (2), except that extreme points belonging to the primitive surfaces of the object will be listed, requiring additional computation using transformation matrices S, R, and T.

(4) Listing the coordinates (x_{mj}, y_{mj}, z_{mj}) of n extreme points, where $j = 1, 2, \ldots, n$, which envelop the object itself but not its bounding space, from which the bounding-space extreme points will be computed. The (x_{mj}, y_{mj}, z_{mj}) coordinates will either be entered manually or be computed from instances pointed out by an operator of an interactive-system display unit.

The steps of computing the bounding window on the basis of the acquisition element and the bounding space are as follows:

(1) Having acquired an element that matches any of the regions of the object to be found, identify the acquisition element with Point 0_b and select the corresponding extreme points on the basis of the element attributes (orientation and color) and the direction (upward or downward) of the acquisition scanning.

(2) Compute the global (x, y, z) coordinates of the acquisition element.

(3) Determine the rotation and translation transformation of the bounding-space coordinates (x_b, y_b, z_b) into the global coordinates (x, y, z).

(4) Using the results of Eqs. (1) through (3), compute the global (x, y, z) coordinates of the remaining extreme points.

(5) Apply the (x, y, z) values of each extreme point into Eqs. (14) and (15) and obtain the image coordinates (x_p, z_p) of each vertex of the polygonal window. Convert the (x_p, z_p) coordinates into (I, J) coordinates.

4.2 Bounding Space for a Single Acquisition Element

Given a single acquisition element of an object, the corresponding bounding space is obtained by assuming that the acquisition element may be located at one of two extreme edge points on the object surface. We distinguish between the cases in which the acquisition element belongs to a vertical surface and a horizontal surface because the bounding space for a vertical element is different from that of a horizontal element.

An additional consideration comparing the acquisition of horizontal and vertical surfaces is as follows. A horizontal region in an office environment is likely to be supporting other objects that yield color and surface attributes different from those of the horizontal region. Although a vertical region may also be occluded by objects in the scene, a larger portion of its area is likely to be exposed compared with that of a horizontal region. Consequently, the acquisition (and validation) of a vertical region is likely to be easier and more reliable (it will have a higher figure of merit) than that of a horizontal region. Fortunately, the upward scanning to be employed in the acquisition of office objects below the horizon will hit vertical regions first for most objects.

4.2.1 Vertical Acquisition Element

Consider in Figure 6(a) the front view of a vertical region of an object model that can be rotated only around the z axis, e.g., standing on the floor. Suppose that a vertical element that matches the color of this region has been acquired. Assuming that the acquisition element is an edge point, four edge points, marked by 1, 2, 3, and 4 in

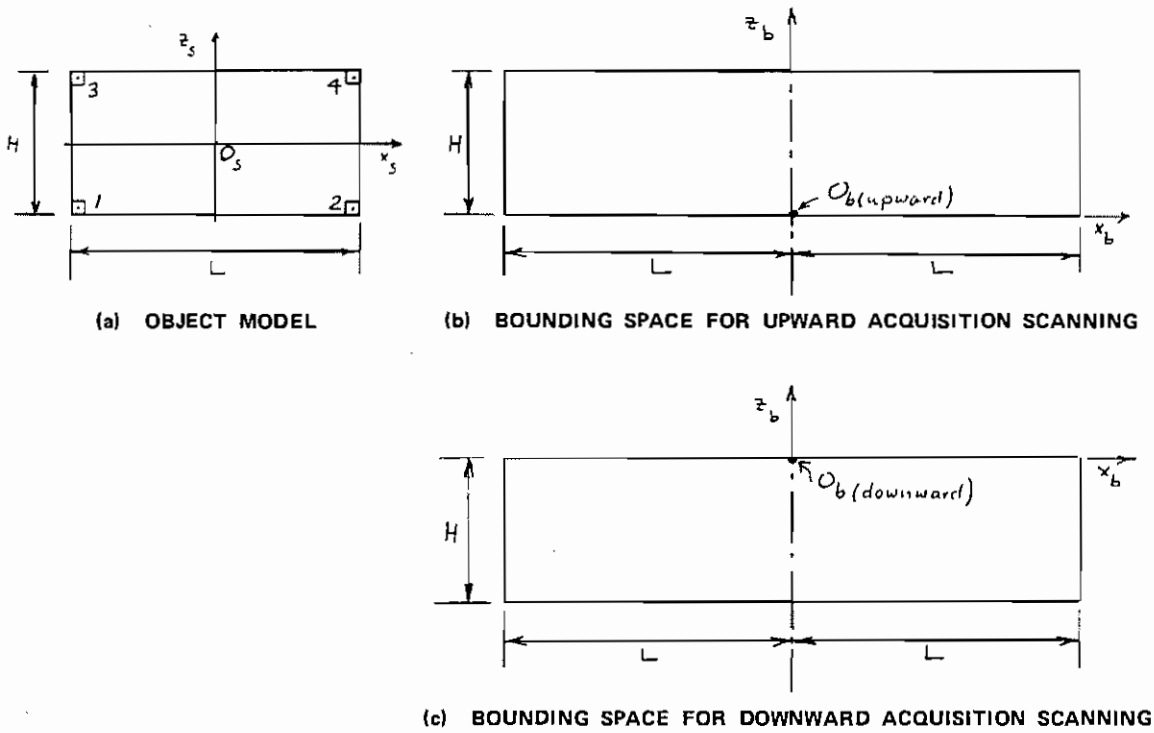


FIGURE 6 FRONT VIEWS OF VERTICAL REGIONS OF AN OBJECT MODEL AND ITS UPWARD-SCANNING AND DOWNWARD-SCANNING BOUNDING SPACES

Figure 6(a), are considered. In determining the bounding space, two cases are distinguished:

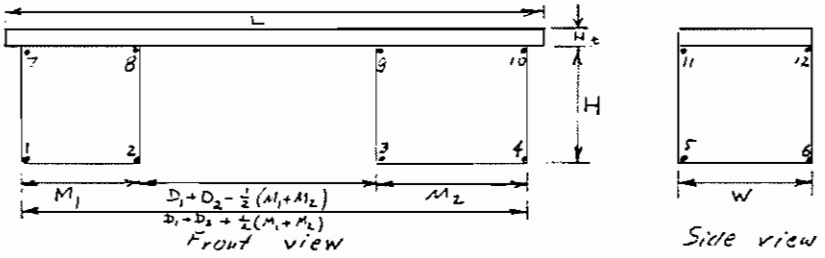
(1) Upward acquisition scanning (decreasing I values), in which the bottom points (Points 1 and 2) will be acquired first.

(2) Downward acquisition scanning (increasing I values), in which the top points (Points 3 and 4) will be acquired first.

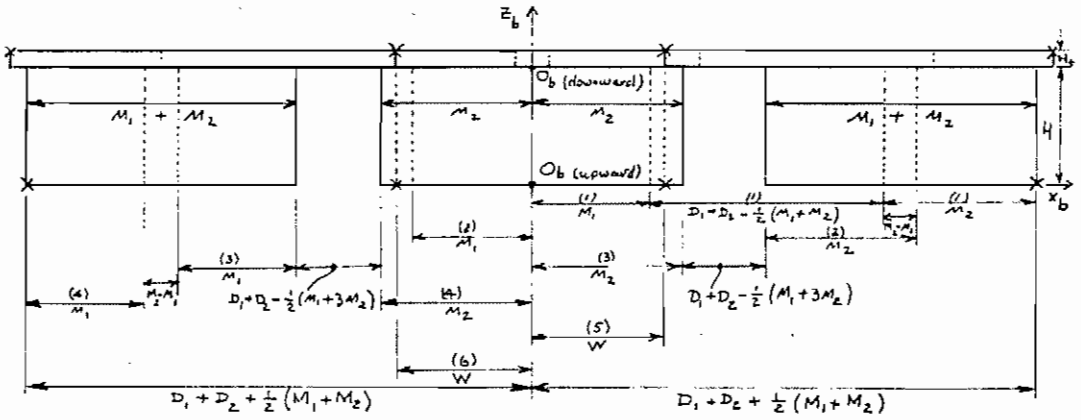
Hence, if the scanning is upward, the extreme acquisition element could be either Point 1 or Point 2. The resulting bounding space is obtained by sliding the object region by its length L, as shown in Figure 6(b). On the other hand, if the scanning is downward, the extreme acquisition element could be either Point 3 or Point 4, and the resulting bounding space is shown in Figure 6(c). The difference between the two bounding spaces is the location of the origin, Point O_b.

In conclusion, the bounding space for a vertical acquisition element is obtained by sliding the front view of the corresponding vertical region along its constrained direction by an amount equal to the length of the region in that direction. The origin O_b is at the bottom of the bounding space for upward scanning and at the top for downward scanning.

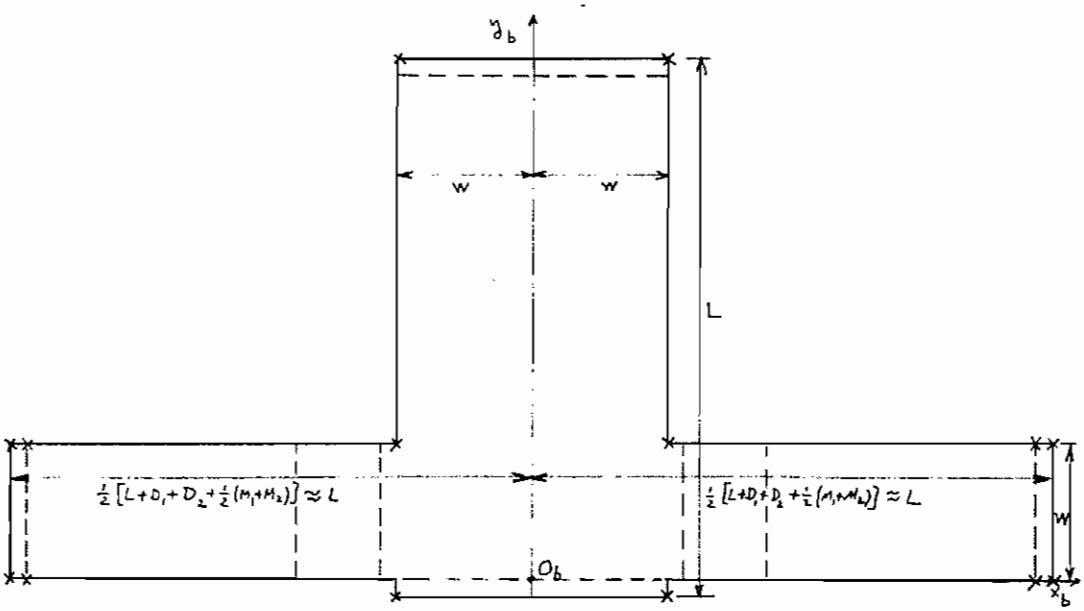
As an example, suppose that we have to find a desk, whose model is shown in Figure 4(a), and that a vertical element has been acquired. Suppose also that the acquisition element could belong to either the front view or the side view. These views are shown in Figure 7(a). For an upward scanning, the acquisition point may be identified with either one of Points 1 through 4 in the front view or Points 5 and 6 in the side view. For a downward scanning, the same is true for Points 7 through 10 in the front view or Points 11 and 12 in the side view. Assuming an upward scanning, the front view of the bounding space is



(a) GEOMETRICAL MODEL OF DESK

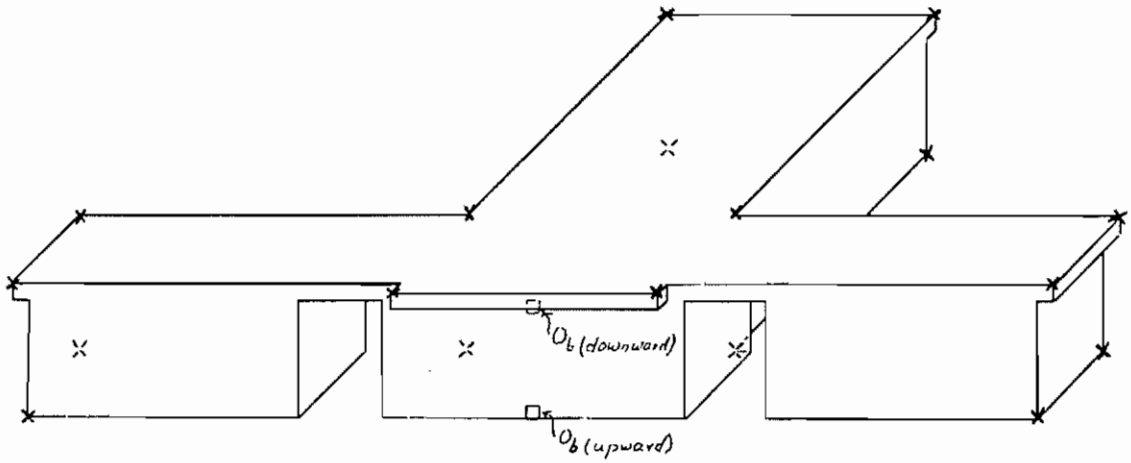


(b) FRONT VIEW OF BOUNDING SPACE



(c) TOP VIEW OF BOUNDING SPACE

FIGURE 7 BOUNDING SPACE FOR A VERTICAL ACQUISITION ELEMENT OF A DESK



(d) PERSPECTIVE VIEW OF BOUNDING SPACE

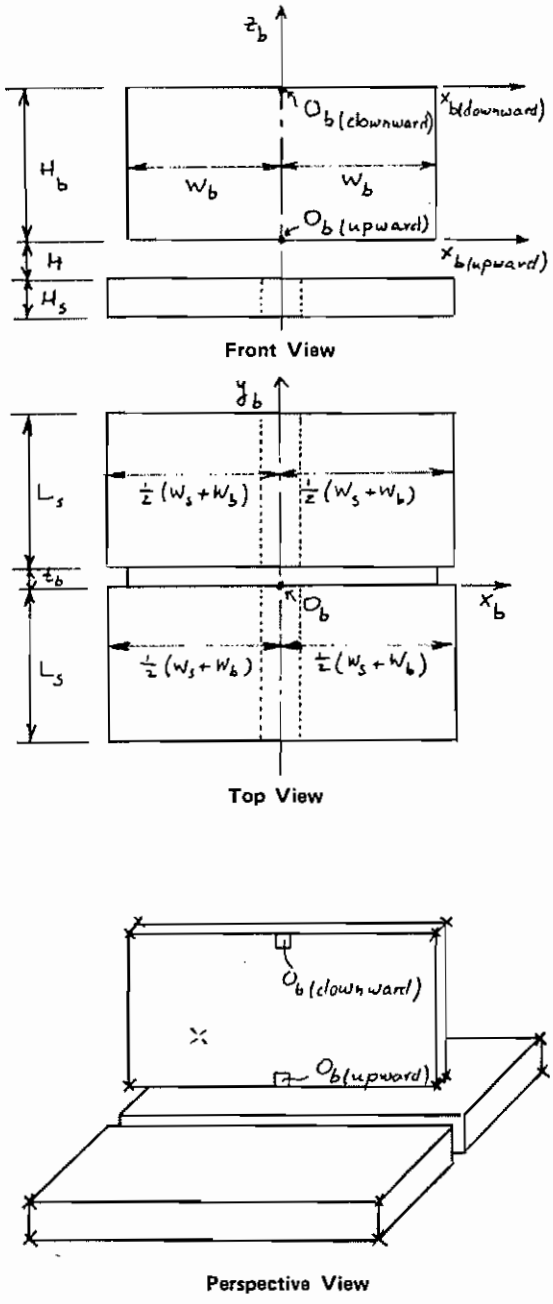
Extreme points are marked by x if seen; by ⁜ if hidden. $n_x = 18$.
Acquisition centroid is identified with Point O_b (upward) for upward acquisition scanning and with Point O_b (downward) for downward acquisition scanning.

FIGURE 7 (Concluded)

constructed in Figure 7(b) by sliding the front view of the desk as each of Points 1 through 6 is identified with the centroid of the acquisition element, Point $0_{b(upward)}$. The corresponding top view is shown in Figure 7(c). Note that the front and top views are symmetrical with respect to the z_b and y_b axes, respectively. Eighteen extreme points that envelop the bounding space are marked by x's. A perspective view of the bounding space, including the 18 extreme points, is shown in Figure 7(d). Note that the two empty spaces under the top of the bounding space are not enveloped by extreme points because they are relatively small and the resulting n_x value is smaller. The (x_b, y_b, z_b) coordinates of each of the 18 extreme points are stored in the bounding table.

Suppose now that the acquisition of a front view of a desk is distinguishable from that of a side view (owing, perhaps, to different colors or/and surface orientation relative to the wall). In such a case the bounding space will be reduced to the corresponding portion of the one shown in Figure 7. Specifically, it will be approximately a rectangular prism: 2L X W X H for the front view, and 2W X L X H for the side view.

For some objects, the acquisition of a vertical element may entail a front-rear ambiguity. For example, a vertical acquisition element of a chair may belong to either the front or the rear of the back support. Based on the chair model in Figure 4(b), the resulting bounding space, including the extreme points, is shown in Figure 8(a). On the other hand, if the vertical acquisition element matches only the front of the back support, then the bounding space is smaller, as shown in Figure 8(b). For either case there are $n_x = 16$ extreme points, as shown by the perspective views in Figures 8(a) and 8(b). If the thickness t_b of the back support is negligible, then 4 extreme points may be eliminated, resulting in $n_x = 12$. If, in addition, the thickness H_s of the seat is ignored, then another 4 extreme points may be deleted, and $n_x = 8$.



(a) ACQUISITION ELEMENT BELONGING TO EITHER FRONT OR REAR OF BACK SUPPORT. $n_x = 16$

(b) ACQUISITION ELEMENT BELONGING TO FRONT OF BACK SUPPORT. $n_x = 16$

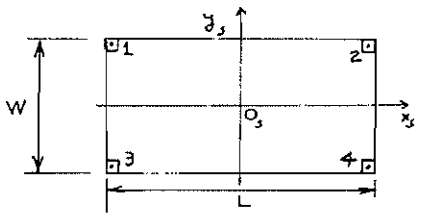
FIGURE 8 BOUNDING SPACE FOR A VERTICAL ELEMENT OF A CHAIR

4.2.2 Horizontal Acquisition Element

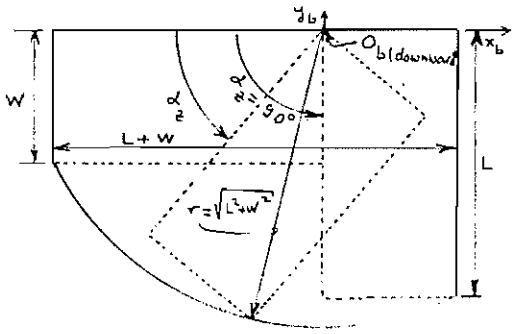
The top view of a horizontal surface of an object model is shown in Figure 9(a). Either one of four edge points, marked 1 through 4, may be identified with the centroid of the acquisition element. For an object consisting of a vertical region below the horizontal region (e.g., a desk), the horizontal region will most likely be acquired by a downward scanning, and the acquisition element will be identified with either Point 1 or Point 2. The converse is true for an object consisting of a vertical region above the horizontal region (e.g., a chair), in which case the acquisition element will be identified with either Point 3 or Point 4.

Consider first downward scanning. If Point 2 has been identified with the acquisition element, then the bounding space is as shown in Figure 9(b). It is obtained by revolving the top view of the object model counterclockwise around the z_b axis by the angle α_z, where $0 \leq \alpha_z \leq 90^\circ$. Similarly, if Point 1 has been identified with the acquisition element, then the bounding space is obtained by 90° clockwise rotation, as shown in Figure 9(c). Now, if the acquisition element may be identified with either Point 1 or Point 2, then the bounding space is obtained by aligning the above two bounding spaces so that their x_b and y_b axes overlap and by finding their union, as shown in Figure 9(d). Using the same reasons, the top view of the bounding space for upward scanning is as shown in Figure 9(e).

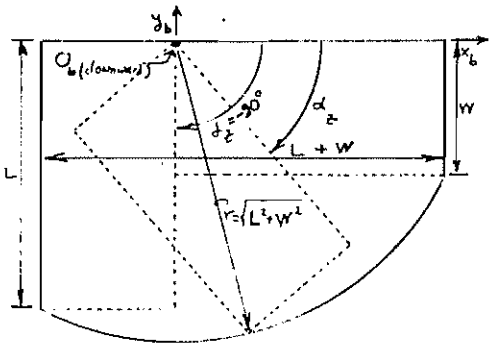
In conclusion, the bounding space for a horizontal acquisition element is the union of the spaces obtained by revolving the top view of the corresponding horizontal region around each of the two edge points by $+90^\circ$ and -90°. The top view of the resulting bounding space is a sectioned semicircle whose origin 0_b is at the top for downward scanning and at the bottom for upward scanning. The bounding space itself consists



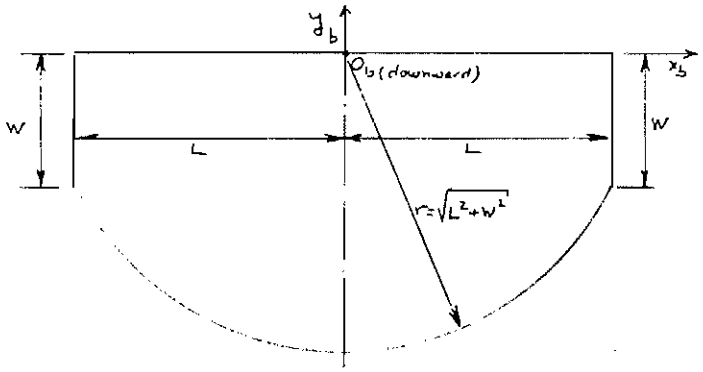
(a) OBJECT MODEL



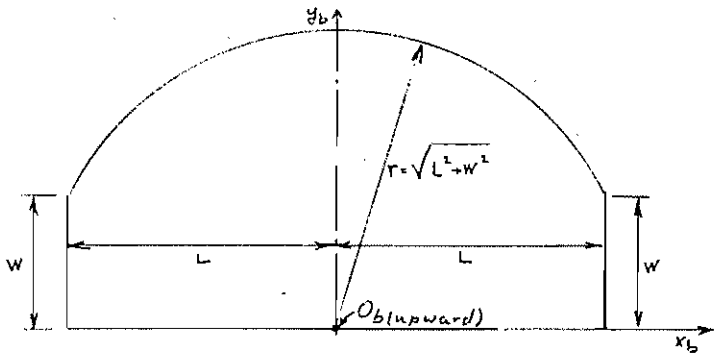
(b) BOUNDING SPACE FOR ACQUISITION ELEMENT 2



(c) BOUNDING SPACE FOR ACQUISITION ELEMENT 1



(d) BOUNDING SPACE FOR DOWNWARD ACQUISITION SCANNING



(e) BOUNDING SPACE FOR UPWARD ACQUISITION SCANNING

FIGURE 9 TOP VIEWS OF HORIZONTAL REGIONS OF AN OBJECT MODEL AND ITS DIFFERENT BOUNDING SPACES

of semicylinders, which may be sectioned or unsectioned, solid or hollow, depending on the rest of the object model.

An example is shown in Figure 10 for a horizontal acquisition element, obtained by downward scanning, that matches the top of a desk. The asymmetry in the sectioned semicylinders stems from the asymmetry in the desk model, Figure 4(a). Eight extreme points envelop the sectioned semicircular top. Four of these are obtained by dividing the circular portion of the top into eight equal angles, each of magnitude γ, as shown in Figure 10. Eight additional extreme points are obtained in a like manner for the bottom of the semicylinder (of height H). Totally, therefore, $n = 16$; however, the four extreme points enveloping the circular portion of the desk top may be eliminated because their images are likely to be inside the bounding window obtained without them, resulting in $n_x = 12$.

Consider now a horizontal acquisition element of a chair whose model is given in Figure 4(b). The bounding space consists of the union of three semicylinders:

(1) A solid, sectioned semicylinder, similar to that of the desk top in Figure 10, which bounds the seat. Depending on the scanning direction, its top view is similar to the one shown in either Figure 9(d) or Figure 9(e).

(2) A hollow semicylinder obtained by revolving the back support around the z_b axis under the assumption that the seat is nearer to the camera, i.e., the front of the support can be seen.

(3) A hollow semicylinder similar to the one in Item 2, except that the seat is farther from the camera, i.e., the rear of the support can be seen.

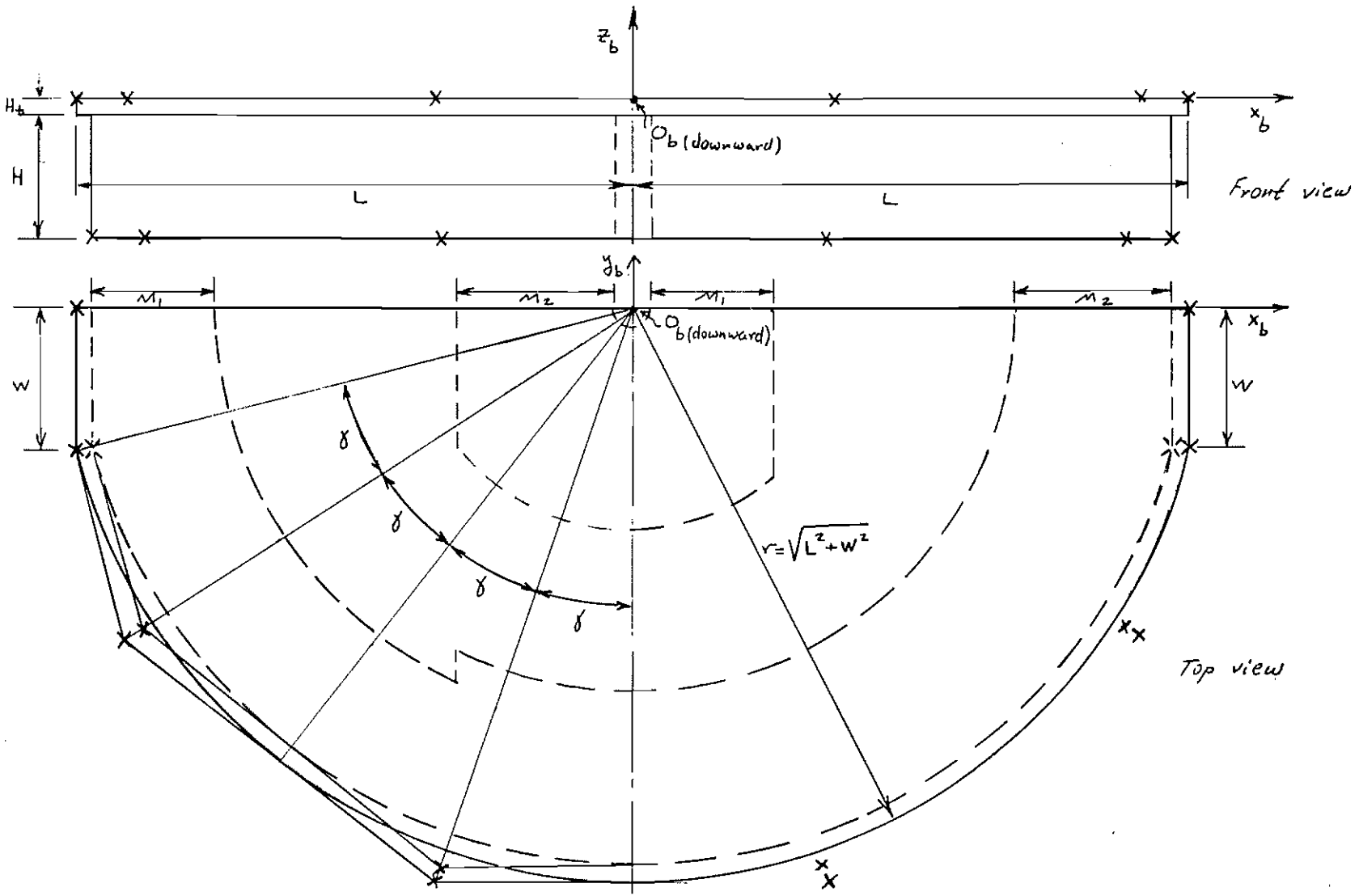


FIGURE 10 BOUNDING SPACE FOR A HORIZONTAL ACQUISITION ELEMENT OF A DESK

Extreme points are marked by x's.

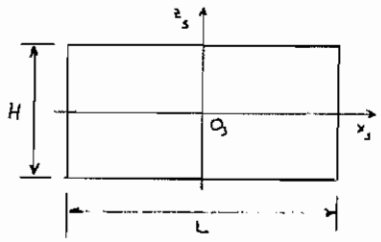
4.3 Bounding Space for Multiple Acquisition Elements

Given more than one acquisition element, we wish to determine the corresponding bounding space. Clearly, with more information than with a single acquisition element, the bounding space will be smaller. The problem is by how much and whether the computation entailed justifies the reduction in the final bounding window. As in the case of a single acquisition element, we shall distinguish between sets of vertical and horizontal acquisition elements, each belonging to the same region. It turns out that the extension of the bounding space from a single acquisition element to multiple acquisition elements is straightforward and easy for vertical elements but is not easy for horizontal elements. Finally, there is the problem of determining the bounding subwindows of individual regions of an object model on the basis of multiple acquisition elements.

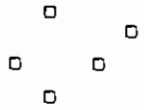
4.3.1 Vertical Acquisition Elements

Consider the front view of a vertical object region of length L and height H in Figure 11(a), and suppose that the multiple acquisition elements shown in Figure 11(b) match this region. To find the bounding space, these multiple elements are first enclosed by a minimal rectangular frame of length ℓ_F and height h_F, to be called "acquisition frame," as shown in Figure 11(c), where $\ell_F \leq L$ and $h_F \leq H$. The z_b axis is arbitrarily identified with the left side of the acquisition frame. The x_b axis is identified with either the bottom or the top of the acquisition frame, depending on whether the scanning is upward or downward, respectively.

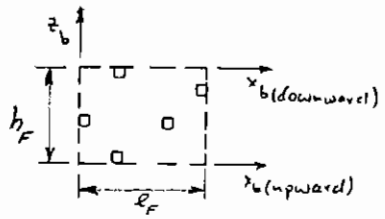
For upward scanning, the acquisition frame overlaps the bottom-left corner and the bottom-right corner of the object model, as shown in Figure 11(d). The length of the resulting space is $2L - \ell_F$.



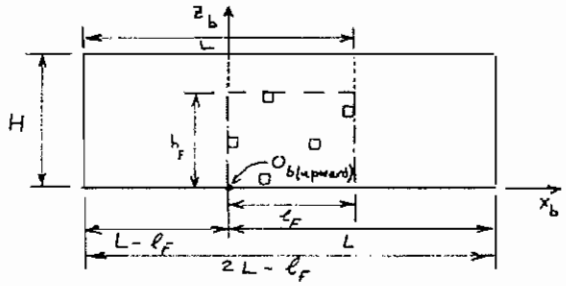
(a) OBJECT MODEL



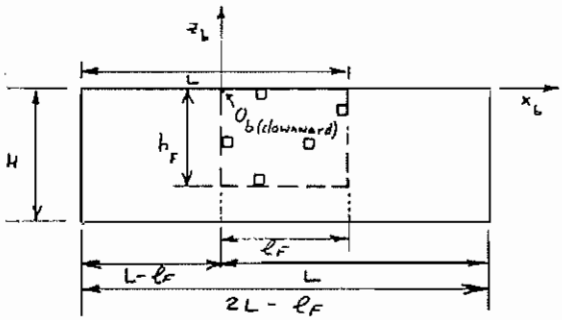
(b) MULTIPLE ACQUISITION ELEMENTS



(c) ACQUISITION FRAME



(d) BOUNDING SPACE FOR UPWARD ACQUISITION SCANNING



(e) BOUNDING SPACE FOR DOWNWARD ACQUISITION SCANNING

FIGURE 11 FRONT VIEWS OF VERTICAL REGION OF AN OBJECT MODEL, ACQUISITION FRAME, AND THE BOUNDING SPACES FOR MULTIPLE ACQUISITION ELEMENTS

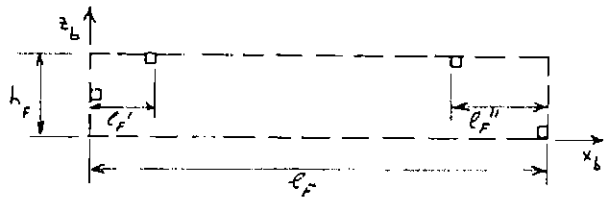
Note in Figure 6(b) that the length of the bounding space for a single acquisition element is 2L, which is a special case in which $\ell_F = 0$.

For downward scanning the acquisition frame overlaps the top left corner and the top right corner of the object model, as shown in Figure 11(e). The length of the bounding space is also $2L - \ell_F$.

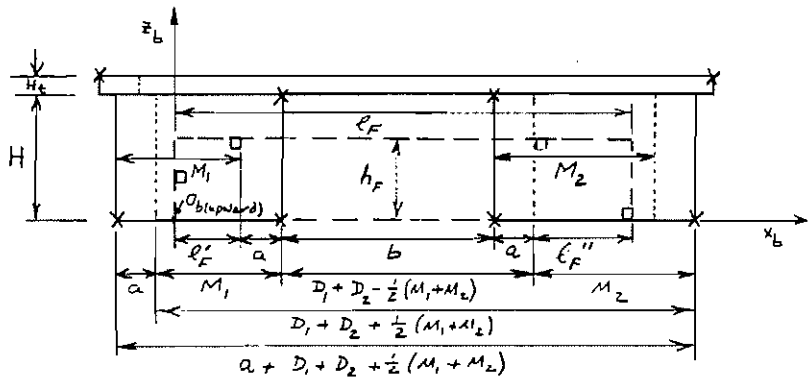
As an example, suppose that again a desk is sought and that four matching vertical acquisition elements, shown in Figure 12(a), have been obtained by upward scanning. The color and surface orientation of each of these elements match either the front view or the side view of the desk model, Figure 4(a). Next, we frame the acquisition elements, Figure 12(b), and find that $W < \ell_F < D_1 + D_2 + 1/2\ (M_1 + M_2)$. We therefore rule out the side view and associate the acquisition elements with the front view. We now slide the front view of the desk model over the acquisition frame by the maximum amount while keeping the acquisition elements within the regions of the front view, as shown in Figure 12(c). The front and top view of the resulting bounding space are shown in Figures 12(c) and 12(d), respectively. The bounding space is enveloped by $n_x = 12$ extreme points, marked by x's.

The bounding space for a desk in Figure 12 is so much smaller than the one in Figure 7 that there is a temptation to look for the subwindows associated with each region of the desk model. By sliding the front view over the acquisition frame as shown in Figure 12(c), each of the desk regions moves between two extreme positions. The union of the images of these positions is the desired region subwindow. For a small bounding space, like that in Figure 12, the overlap among the different region subwindows is much smaller than the one in Figure 7; it is similar to the one assumed in Figure 2(f).

(a) ACQUISITION ELEMENTS OBTAINED BY UPWARD SCANNING



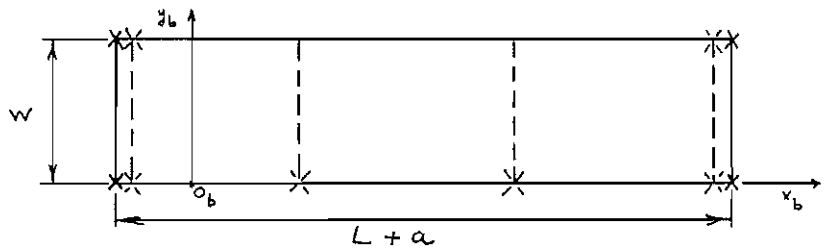
(b) ACQUISITION FRAME



$$a = \ell_F - \left[\ell_F' + \ell_F'' + D_1 + D_2 - \tfrac{1}{2}(M_1 + M_2)\right]$$

$$b = 2(D_1+D_2) - (M_1+M_2) + \ell_F' + \ell_F'' - \ell_F$$

(c) FRONT VIEW OF BOUNOING SPACE



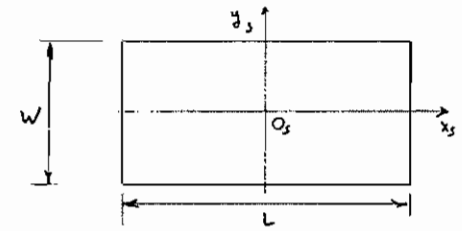
(d) TOP VIEW OF BOUNDING SPACE. $n_x = 12$

FIGURE 12 BOUNDING SPACE FOR A MULTIPLE OF VERTICAL ACQUISITION ELEMENTS BELONGING TO THE FRONT REGION OF A DESK

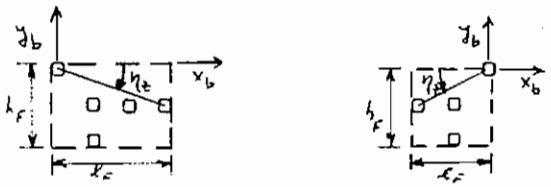
4.3.2 Horizontal Acquisition Elements

The top view of a horizontal surface of an object model is shown in Figure 13(a). Suppose that a multiple of horizontal acquisition elements that match the object model has been acquired during downward scanning. Two different sets of multiple acquisition elements are shown in Figure 13(b). Each set is enclosed by an $\ell_F \times h_F$ acquisition frame. For each set the highest element is selected as the origin of the (x_b, y_b) axes, and the angle between the x_b axis and the straight line connecting the highest and the next highest elements is denoted by η_z. For the set of acquisition elements in Figure 13 (b-i), $\eta_z < 0$, but for the one in Figure 13 (b-ii), $\eta_z > 0$. The top view of the bounding space for each of these sets is shown in Figure 13(c), and its construction is explained as follows.

Consider first the acquisition set in Figure 13 (b-i), for which $\eta_z < 0$. In Figure 13 (c-i) the object model is first laid on top of the acquisition frame with the top left corners overlapping. The object model is then slid to the left by the maximum amount without losing the rightmost acquisition element. The object model is then revolved around Point $O_{b(\text{downward})}$ by an angle varying from 0 to η_z. For each of the revolved positions, the object model is slid by the maximum amount without losing the rightmost acquisition element to take care of the possibility that a portion of the object model has not been acquired because of occlusion. The object model is now positioned so that its edge of length L is along the y_b axis. From that position it is revolved as before around Point $O_{b(\text{downward})}$ by an angle varying from 0 to η_z. The resulting bounding space has an odd-looking shape, as shown by the solid line in Figure 13 (c-i).

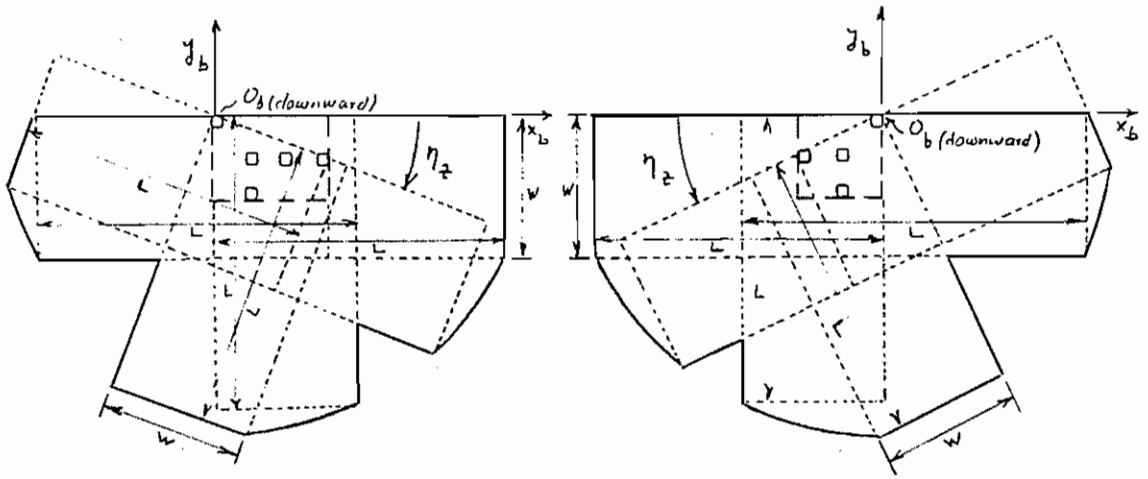


(a) OBJECT MODEL



i. $\eta_z < 0$ ii. $\eta_z > 0$

(b) ACQUISITION FRAME



i. $\eta_z < 0$ ii. $\eta_z > 0$

(c) BOUNDING SPACE FOR DOWNWARO SCANNING

FIGURE 13 TOP VIEWS OF HORIZONTAL REGION AND BOUNDING SPACES FOR DIFFERENT MULTIPLES OF HORIZONTAL ACQUISITION ELEMENTS

The bounding space for the acquisition set shown in Figure 13 (b-ii), for which $\eta_z > 0$, is constructed in a similar manner except that the left and the right shapes are interchanged. The resulting bounding space is shown in Figure 13 (c-ii).

ACKNOWLEDGMENTS

The strategy illustrated in Section 1 is based on goal-directed perceptual strategies developed by Marty Tenenbaum and Tom Garvey. The modules described in Sections 3 and 4 were developed in response to the requirements of these strategies.

REFERENCES

1. J. M. Tenenbaum, "On Locating Objects by Their Distinguishing Features in Multisensory Images," Artificial Intelligence Center, Technical Note 84, SRI Project 1187 and 1530, Stanford Research Institute, Menlo Park, California (September 1973).

2. D. Nitzan, "Scene Analysis Using Range Data," Artificial Intelligence Center, Technical Note 69, SRI Project 1530, Stanford Research Institute, Menlo Park, California (August 1972).

3. R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis" (John Wiley & Sons, New York, New York, 1973).