# Generic Object Recognition Using CAD-Based Multiple Representations

*Subhodev Das* and *Bir Bhanu*
College of Engineering
University of California
Riverside, CA 92521-0425

*Chih-Cheng Ho*
Department of Computer Science
University of Utah
Salt Lake City, UT 84112

## Abstract

Real-world applications of computer vision usually involve a variety of object models making a single model representation somewhat inadequate for object recognition. Multiple representations, on the other hand, allow different matching strategies to be applied for the same object, or even for different parts of the same object. This paper is concerned with the use of CAD-derived hierarchical models having multiple representations – concave/convex edges and straight homogeneous generalized cylinder – for generic object recognition in outdoor visible imagery. It also presents a refocused matching algorithm that uses a hierarchically structured model database to facilitate generic object recognition. Experimental results demonstrating generic recognition of objects in perspective, aerial images are presented.

## 1 INTRODUCTION

The success of a three-dimensional (3-D) model-based object recognition scheme is dependent on several factors: representation of the model (e.g., wire-frame, constructive solid geometry, surface boundary or B-rep), types of features in the input image, choice of search technique to match the model and the data. In real-world image understanding (IU) problems, there are additional factors that complicate the overall model-based object recognition process. These include occlusion, shadow, cloud cover, haze, seasonal variations, clutter, and various other forms of image degradation. Typically, CAD models of objects are used in IU tasks involving man-made objects [2]. However, most IU systems use a single representation for the models or a matching technique based on a single representation for object recognition. On the other hand, for real-world applications of IU, it is unlikely that a single representation-based recognition strategy would suffice for a variety of objects; multiple representations allowing different matching strategies to be applied for the same object, or even for different parts of the same object, is a better alternative that has not been explored in the past.

This paper is concerned with (a) building appropriately detailed, hierarchical models of objects with multiple representations and (b) using these models to perform generic object recognition in outdoor imagery. From a computational point of view, hierarchical representation simplifies the complexity of the problem and provides a solution for recognition of partially visible objects even under self occlusion. Moreover, psychological studies [3, 7] have given evidence of the role of parts in human visual recognition. Apparently, multiple representations are derived for each subpart and are used to guide the extraction of image features. During the matching phase, these representations admit cooperative matching strategies and simultaneous verification of the subparts of a hypothesized object. Generic object recognition is the process of determining the class of objects that is most *similar in shape* to the actual image and is accomplished using shape models of *unknown* dimensions. The generic object recognition systems proposed in the past [1, 4, 8] are either simplistic or inadequate for handling shadows, clutter and other image degradations encountered in complex image interpretation tasks. The imaging conditions and the viewpoint location can affect the sensory data in such a way that their validation using detailed CAD models may be difficult. Thus, this paper also emphasizes the use of approximations of CAD models for generic object recognition.

In the following sections, we discuss the details of the generic object recognition approach that utilizes hierarchical CAD models with multiple representations. Section 2 describes creating appropriately detailed, hierarchical vision models with multiple representations. Section 3 discusses the recognition algorithm using these hierarchical models. Section 4 gives the details of implementation and the experimental results using real-world data. Section 5 presents the concluding remarks.



0-8186-5310-8/94 $03.00 

## 2 BUILDING GENERIC OBJECT MODELS WITH MULTIPLE REPRESENTATIONS

We begin with a discussion on the need for view-dependent model descriptions and derivation of the conditions, as functions of the imaging geometry, under which these descriptions may be obtained for given CAD models.

### 2.1 Viewpoint-dependent Representation

There is a need for an adequate model description, even for generic CAD models, which can be supported by the sensory data. We, therefore, emphasize the use of appropriately detailed models (for the purpose of model-image matching), to be called the ***diffused models***, where the degree of detail is determined by the viewing conditions or the imaging parameters.

Consider the perspective projection of a planar curve, $C$, as illustrated in Figure 1. Let $\mathbf{X}_1 = (X_1, Y_1, Z_0)$ and $\mathbf{X}_2 = (X_2, Y_2, Z_0)$ denote two points on $C$ whose image projections are $(r_1, c_1)$ and $(r_2, c_2)$, respectively. According to Figure 1,

$$\Delta r = \frac{fk_y}{Z_0}|X_2 - X_1| \quad \text{and} \quad \Delta c = \frac{fk_x}{Z_0}|Y_2 - Y_1|, \quad (1)$$

where $\Delta r = |r_2 - r_1|$ and $\Delta c = |c_2 - c_1|$, $f$ is the focal length of the imaging system, and $k_x$ and $k_y$ are internal parameters of the sensor. In order for $\mathbf{X}_1$ and $\mathbf{X}_2$ to be distinct, $\Delta r \geq 1$ pixel and $\Delta c \geq 1$ pixel, i.e., the resolution at the plane must be

$$\Delta X = |X_2 - X_1| \geq \frac{Z_0}{fk_y} \quad \text{and} \quad \Delta Y = |Y_2 - Y_1| \geq \frac{Z_0}{fk_x}. \quad (2)$$

A diffused model of $C$ for the given resolution $(\Delta X, \Delta Y)$ is such that the two shapes (due to the original and diffused models) are equivalent.

To compare two shape models, we utilize the notion of "$\epsilon$-neighborhood" of a point ([6], a spherical region of radius $\epsilon$ centered at that point). The "distance" between two shapes is the smallest value of $\epsilon$ for which each shape is *completely contained* within the spherical neighborhood of the other, perhaps after some arbitrary displacement.

**Proposition 1** *Two shapes are said to be* equivalent *if their distance is within the limit of the specified resolution.*

In our case, the smallest $\epsilon$ corresponding to the distance between $C$ and its diffused model is $\epsilon = \sqrt{(\Delta X)^2 + (\Delta Y)^2}$. This tolerance can be used to derive diffused models of objects as discussed next.

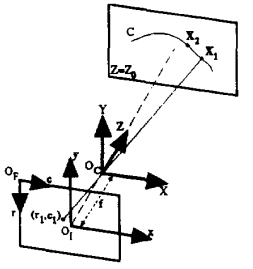


Figure 1: Perspective projection of a planar curve.

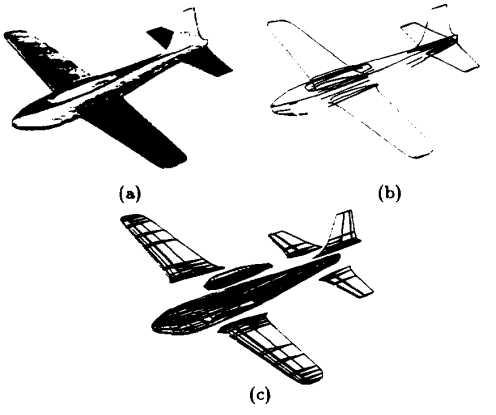
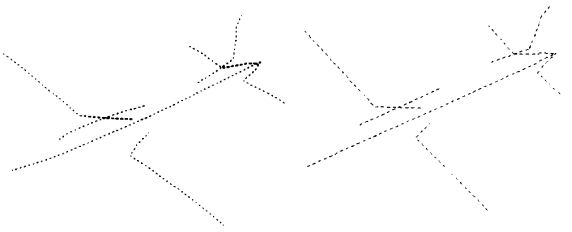



Figure 2: A generic airplane model: (a) B-spline model of an airplane, (b) Edge detection using a polyhedral approximation of the B-spline model, (c) Decomposition of the polyhedral approximation along concave edges.

### 2.2 Diffused Models with Multiple Representations

The CAD system used in this work is the Alpha_1 solid modeling system developed at the University of Utah. Figure 2(a) shows an airplane CAD model whose decomposition is shown in Figure 2(c). The decomposition is obtained by first deriving a polyhedral approximation of the B-spline model of Figure 2(a), and then finding the concave discontinuities of surface normals (Figure 2(b)) that separate the different subparts. The final object models are represented by constructing relational links between their respective part decompositions. The details of the approach appear in Bhanu and Ho [2].

Given a space curve and a diffusion limit $\epsilon$ based on resolution it is possible to obtain a (nonunique) dif-



(a) (b)

Figure 3: Generalized cylinder representation of subparts of an airplane: (a) Axes of GCs, (b) Axes of SHGC approximations of GCs in (a).

fused model of the curve by incrementally deforming the original curve to obtain progressively lower-order space curves. At each step of the iterative process, position and orientation of the most recent deformed curve can be calculated with respect to the original curve to verify that the amount of displacement is within the resolution limits. To illustrate this procedure, consider a GC-based description of a CAD model whose axes and cross-sections are typically described by low-order polynomials. For every subpart of a decomposed object model represented using a GC, it is tested whether a straight homogeneous generalized cylinder (SHGC) approximation is feasible for that part, i.e., the axis curve can be approximated by a straight line and all the different cross-sectional functions can be replaced by a single homogeneous function. If the conditions are satisfied, then an SHGC-based representation is obtained for that subpart. Figure 3(a) shows the axes of GC representations of the subparts of the original (non-diffused) airplane model and Figure 3(b) describes the axes of the corresponding subparts of the diffused model using the SHGC representation. The model of Figure 3(b) can be further approximated depending on $\epsilon$-value to obtain a diffused model in which small SHGCs, i.e., ones with axes lengths nearly equal to or smaller than $\epsilon$, are ignored, e.g., the SHGCs connecting the wings or the tails to the fuselage in Figure 3(b).

Multiple-representation descriptions are maintained for each of the decomposed subparts using polyhedral approximation, concave/convex edges, curvature extrema, surface normals, and generalized cylinders. Two examples of multiple representations of object models are shown for the airplane model: the edge (or curvature extrema) points of Figure 2(b), and the generalized cylinder representation of Figure 3.

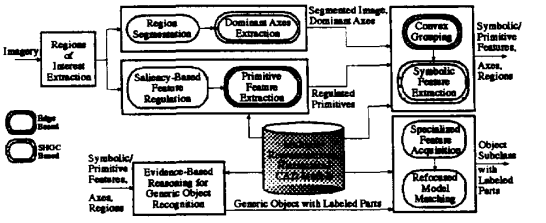



Figure 4: A flow diagram of the *generic* object recognition algorithm.

## 3 RECOGNIZING GENERIC OBJECT MODELS WITH MULTIPLE REPRESENTATIONS

Given multiple representations of hierarchical, diffused CAD models, the goal of object recognition is to utilize these multiple representations of the subparts during the recognition process. The use of multiple representations facilitates cooperative interaction among different matching strategies. The recognition process initiates with the most generic, diffused CAD models. A cycle of increasingly specific feature extraction and finer recognition, called refocused matching, then allows the system to interpret a scene in terms of individual (more specific) CAD models. The algorithm for generic object recognition consists of the steps indicated in Figure 4 that are monitored by a supervisory controller.

The input to the recognition system is a 2-D image and ancillary data about the imaging and platform parameters and scene conditions. Initially, a multiresolution search is performed to determine the regions of interest (ROIs) that are likely to contain the target objects. Once the ROIs have been identified, each of them is subjected to the recognition process described below in succession.

### 3.1 SHGC Representation-Based Dominant Axes Extraction

The dominant axes correspond to the longest axes of the SHGC representations of CAD models. Normally, the contours of SHGCs are used for determining the axes of SHGCs from edge images. However, in outdoor images, where contrast and image quality can vary greatly, a reliable extraction of edges is often not possible. In our approach, the 2-D shapes of objects are used to extract the dominant axes from the images. Such shapes are identified by segmenting image regions that may

correspond to the objects of interest. The segmentation is based on the joint relaxation of a two-class (object/background) region-based approach and a two-class (edge/no edge) edge-based approach.

To extract dominant axes of an SHGC-based representation, we note that the curvature at an end point of an SHGC is inversely proportional to the distance from its origin of scaling. The approach to extracting the potential dominant axes of an SHGC representation-based shape involves identification of the high curvature points of the region boundaries corresponding to the contours of the subparts represented using SHGCs. To determine the high curvature points along the region boundaries, the minimum bounding polygon (MBP), i.e., the smallest (area-wise) convex polygon, that completely encloses the object region is found. It can be shown that the vertices of the MBP lie close to the local extrema of curvature points along the region boundaries.

Now, more than one "extreme" point may be identified within a small neighborhood along the region boundary in the vicinity of a polygon vertex, e.g., when there are multiple local extrema points in a certain segment of the boundary. In that case, nearby "extreme" points are grouped into clusters and the cluster centers are chosen to represent the region extremities. A potential dominant axis is a line that connects two such extreme points that are not the centers of adjacent clusters. Lines whose significant portions are not contained within the segmented region are ignored.

### 3.2 Edge Representation-Based Primitive Feature Extraction

Edge pixels in images are detected by applying multiple thresholds and are thinned to one-pixel width. Next, *long* edge segments that are made up of *high* magnitude edge pixels are found to form linked edge segments through an optimization step. Each set of linked edge segments or a *configuration level* constitutes perceptually salient contours corresponding to one threshold value. These sets are handed over to the primitive feature extraction process in a regulated manner, starting with the top-level configuration (see Figure 4).

The extraction of primitive features involves edge-based representation of shapes. Since this representation is derived from polygonal approximations of B-spline CAD models, the primitive features comprise linear segments. The input to this line extraction algorithm is a set of regulated intensity-edge segments. Now, the model edges intersect in 3-D corners whose projections are 2-D corners. Thus, in addition to lines, our algorithm also detects corners by obtaining gradient and curvature measurements at pixels in the grey scale image.

### 3.3 Symbolic Feature Extraction

Symbolic features may be derived using perceptually grouped primitives. We adopt a two-stage approach for the extraction of symbolic features. First, the line primitives are organized into convex groups using domain-independent perceptual measures. Second, SHGC-based representation of the hierarchical object models are used to extract symbolic features from these convex groups that may correspond to the different subparts of a generic object.

#### 3.3.1 Edge representation-based convex grouping

Polyhedral edges are either convex or concave; subparts modeled using convex edges give rise to convex groups of lines in the images. Initially, groups of lines are formed based on the perceptual measures of proximity and collinearity. The motivation here is that if the elements of a group belong to a region boundary, then the segmentation results would help to determine the interior of the region and hence to verify the "convexity" of the group.

A convexity test is performed for every pair of lines in a selected group. If a line of a selected pair fails the above convexity test, then that line is removed and put in a new group by itself. After all the initial groups have been considered, this process creates the *first* set of convex groups and isolated lines removed during the convexity test. The *second* pass considers whether an isolated line can be put in a convex group based on proximity, collinearity, and convexity.

#### 3.3.2 SHGC representation-based symbolic feature extraction

The high-order symbolic features are obtained as assemblies of the lower-order perceptual groups by accessing the object models. When these models are represented using SHGCs, the symbolic features correspond to the contours of SHGCs. Since the rules for deriving the symbolic features are based on one representation (viz., SHGC) while the primitive features are extracted based on another representation (viz., convex edges), there must exist a transformation from one representation to the other. To facilitate this transformation, we consider one class of SHGC, the linear right SHGC (LRSHGC), whose contour can be represented using a convex group of lines. For generic models such as the aircraft model of Figure 2(a), subparts can be represented

using LRSHGCs for certain degree of diffusion. Consequently, convex groups of lines are used to extract these subparts such as wings, tails, and the rudder sections of the generic aircraft.

### 3.4 Evidence-Based Reasoning for Recognition by Parts

Our approach to reasoning is "exact" or non-monotonic which we shall refer to as *evidence-based*. This particular reasoning method accumulates evidence, i.e., determines the number of positive evidences, in support of the hypothesized generic object. (In this work, we do not consider negative evidences for a hypothesized object.)

Once the symbolic features have been derived, these need to be matched to the generic object model through the evidence accumulation process. It primarily involves verifying the mutual connectedness of the symbolic features that represent the different parts of a generic object. The exact manner in which this verification is to be carried out is specified by the production rules associated with this object model, but in all cases we emphasize simultaneous verification of hypotheses. The combined support of a body of evidence for a hypothesis is the total number of positive evidences that can be found in the input data. The final output are the identified symbolic parts of the generic object.

### 3.5 Refocused Matching

The labeled symbolic parts are now used to direct the image-based search for more localized features that are available at lower levels of the database hierarchy. The symbolic features may have associated qualitative or quantitative information. When the latter is available, it may be used to derive constraints for the subsequent identification of the detected generic object. We now describe a method for utilizing quantitative information in refocused matching. It makes use of the pose computation approach of DeMenthon and Davis [5].

Referring to Figure 5, let the model coordinate axes be centered at a point $M_0$ and let $M_i$ denote any other model point. Also, let $\mathbf{X_0}$ and $\mathbf{X}_i$ denote the coordinates of these points in the sensor-based world coordinate system and $\mathbf{r}_0$ and $\mathbf{r}_i$ denote the same in the frame coordinates, respectively. Then, by using perspective projection relations one obtains

$$\begin{aligned} r_0 - r_i &= \frac{fk_y}{\bar{Z}}(Y_0 - Y_i) = s_y \mathbf{i} \cdot \mathbf{M_0M_i}, \\ c_0 - c_i &= \frac{fk_x}{\bar{Z}}(X_0 - X_i) = s_x \mathbf{j} \cdot \mathbf{M_0M_i}, \end{aligned} \quad (3)$$

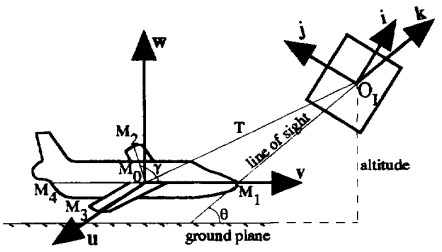



Figure 5: The coordinate systems relating an object model to its projection for pose computation. The vectors in the systems **u-v-w** and **i-j-k** are unit vectors.

where it is assumed that $Z_0 = Z_i = \bar{Z}$ and $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ denote the unit vectors along $X$, $Y$, and $Z$-axes, respectively. The unknown multiplicative factors $s_x = fk_x/\bar{Z}$ and $s_y = fk_y/\bar{Z}$ signify the scaling of the model in $X$- and $Y$-directions, respectively. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ denote the unit vectors along the three axes of the model-centered coordinate system where the sensor and the model unit vector systems are related as $[\mathbf{i}\ \mathbf{j}\ \mathbf{k}]^T = R[\mathbf{u}\ \mathbf{v}\ \mathbf{w}]^T$, $R$ being a $3 \times 3$ orthonormal rotation matrix with rows $[i_u\ i_v\ i_w]$, $[j_u\ j_v\ j_w]$, and $[k_u\ k_v\ k_w]$. For $m$ model points, i.e., $i = 1, ..., m$, the corresponding systems of equations are represented using matrix-vector notation as

$$\mathbf{r} = A\mathbf{I}, \qquad \mathbf{c} = A\mathbf{J}. \quad (4)$$

Here, $A$ is an $m \times 3$ matrix of the coordinates (in the model-based system) of the $m$ model points, $\mathbf{r}$ and $\mathbf{c}$ are $m \times 1$ vectors of the frame coordinates of these points relative to that of $M_0$, and $\mathbf{I} = [s_y i_u\ s_y i_v\ s_y i_w]^T$ and $\mathbf{J} = [s_x j_u\ s_x j_v\ s_x j_w]^T$. If $m = 3$ and the corresponding points are non-coplanar, then $A$ has a full rank and $\mathbf{I}$ and $\mathbf{J}$ can be uniquely solved and hence $\mathbf{i}$ $\mathbf{j}$, $\mathbf{k}$ and $R$. The magnitude of the translation vector is obtained as $|\mathbf{T}| = \sqrt{X_0^2 + Y_0^2 + Z_0^2}$ and its direction is given by the vector $\mathbf{O_I x_0}$. The translation vectors corresponding to all other model points can be similarly obtained. These together with $\mathbf{T}$ allow updating the estimates of $\mathbf{M_0M_i}$'s, which are known only approximately due to the generic nature of the object model, and refining the pose computations in an iterative manner. Once the algorithm has converged, this will yield better estimates of the model parameters than is available initially. These improved estimates can be subsequently used in the identification step.

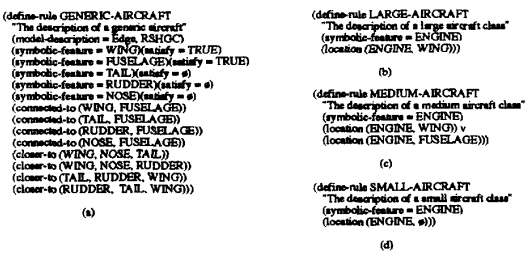



Figure 6: Descriptions of hierarchical CAD models with multiple representations: (a) A generic aircraft, and the three aircraft classes (b) Large, (c) Medium (d) Small.

## 4 EXPERIMENTAL RESULTS

The results reported in this paper are based on one generic object – aircraft – and its three subclasses – large, medium, and small (see Figure 6). Figure 7(a) shows an aerial photograph ($4K \times 4K$) which has several aircraft – four C-130's and one F-18. Using the *multiresolution* focusing approach, several regions of interest are identified as shown in Figures 7(b)-(d) and these are analyzed by the object recognition system in succession. Here, we present the results of analyzing one ROI ($162 \times 240$) from Figure 7 that contains the F-18 aircraft. The ROI and the output of the multi-threshold edge detection step are shown in Figures 8(a)-(c). In our implementation, we have selected five threshold ($t$) values which are fixed for *all* images. The result of extracting globally salient edge contours is presented in Figure 8(d) which shows the top-level configuration, consisting of the aircraft in this case. The following step is to extract the primitive features from this global structure. The result of line fitting to the salient structures is shown in Figures 8(e). Segmented regions and the dominant axes of regions are shown in Figures 8(f). Figure 9(a) shows the six convex sets of lines identified using convex grouping procedure. These are used to extract trapezoid-like features shown in Figure 9(b) that are the symbolic descriptions of some of the subparts, such as wings, tails, and rudder.

During the generic object recognition step, the dominant axes are used to support or refute a selected symbolic feature as a wing of the aircraft or the fuselage. Once all the conditions of connectivity and relative localization of the different subparts have been satisfied, can their ensemble be recognized as a generic aircraft. The identified subparts are shown in Figure 9(c). The connectivity information of the parts is exploited to obtain more complete descriptions of the subparts, followed by the extraction of the shape skeleton. These results are shown in Figures 9(d)-(e). Note that no *precise* model has been utilized in this recognition step. Next, an improved classification of the generic aircraft is sought based on the engine location (Figure 6). However, no elongated blob-like region (symbolic description of an engine) is detected that may indicate presence of engines. Therefore, the generic aircraft is identified as belonging to a *small* class (Figure 9(f)).

The quantitative information associated with the three subclasses of the generic aircraft category is indicated in Table 1. Also available to the system are the location of the sensor, (527, 337, 560) m. in a reference world coordinate system, and the range-to-ground, 805 m, along the line of sight (LOS). The sensor- and model-based coordinate systems for computation of the approximate dimensions of the classified aircraft are shown in Figure 5. The range of $\bar{Z}$ (refer to Section 3.5) measured along the LOS is obtained as 798-801 m based on Table 1 data and Figure 5. The four line segments, indicated as $M_0M_1$, $M_0M_2$, $M_0M_3$, and $M_0M_4$ in Figure 5, of the small aircraft model which are to be used in computing $R$ are further described in Table 2. The $A$ matrix of Eq. (4) consisting of the coordinates of the model points $M_1$, $M_2$, $M_3$, $M_4$ relative to $M_0$ is

$$A = \begin{bmatrix} 0 & a & 0 \\ b\sin\gamma & -b\cos\gamma & d \\ b\sin\gamma & -b\cos\gamma & d \\ 0 & -c & 0 \end{bmatrix},$$

where $3.5 \leq a \leq 11$ m, $5.0 \leq b \leq 11.8$ m, $2.7 \leq c \leq 6$ m, $0 < d < 1.0$ m and $93° \leq \gamma \leq 115°$. To compute $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, the model line $M_0M_4$ is not used since it is collinear with $M_0M_1$. Thus, $A$ is actually a $3 \times 3$ matrix without the last row. The goodness of the computed rotation matrix is evaluated by obtaining the following score: $G = |\mathbf{i} \cdot \mathbf{i} - \mathbf{j} \cdot \mathbf{j}| + |\mathbf{i} \cdot \mathbf{j}|$. The objective here is to identify the range of the 4-tuples $a, b, d, \gamma$ for which $G$ is the lowest. In this experiment, a value of $d = 0.5$ m is selected, which denotes the height of the wings above the **u-v** plane (since this information is not usually available), and the space of $a - b - \gamma$ is searched. The most significant mode of $G$ is observed around 0.6 (an interval size of 0.1 is used to partition $G \in [0, 1]$). It is observed that at least one of $a$, $b$, and $\gamma$ values corresponding to the interval of $G$ around this mode agrees with the groundtruth, i.e., the dimensions of an F-18 aircraft. The ranges of $a$, $b$, and $\gamma$ values for this interval are observed as 8-11 m, 6-6.5 m, and $95° - 115°$, respectively. Some of the triplets belonging to this interval are listed in Table 3. Triplets with lower $G$ values that

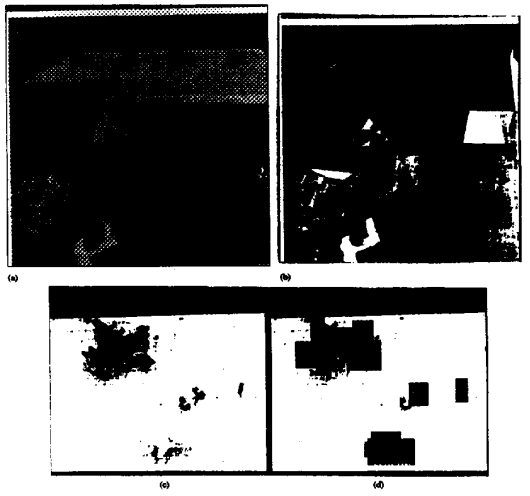


Figure 7: An aerial view of an airfield: (a) Original image ($4K \times 4K$), (b) Preliminary regions of interest (ROIs, black regions) in (a), (c) A close-up of the preliminary ROIs of (b), (d) New ROIs found in (c).

are not associated with any significant mode are also included in Table 3 for comparison. The range values are more precise than the ones listed in Table 1 for the small aircraft class and should facilitate more focused search for the specific object models.

## 5 CONCLUSIONS

In this paper, we have presented an approach to generic recognition of objects using hierarchical, diffused CAD models with multiple representations. Our approach has been based on identifying the hierarchical subparts of a generic object model in the input image and verifying their spatial ordering through an evidence-based reasoning process. The novel aspect of our work is the use of multiple representations for the CAD models of generic objects and the derivation of features and constraints for recognition based on such representations. The generic object recognition strategy emphasized in this paper can serve as the initial step for any model-based object recognition technique.

Table 1: Approximate dimensions of the three subclasses of a generic aircraft.

| Feature | Value | | |
|---|---|---|---|
| | *Large* | *Medium* | *Small* |
| Wingspan | 136'-251' | 66'-107' | 30'-70' |
| Wing Sweep, Leading | 112°-131° | 114°-121° | 93°-115° |
| Wing Sweep, Trailing | 97°-116° | 99°-104° | 84°-100° |
| Fuselage Length | 126'-228' | 52'-91' | 30'-56' |
| Length, Wing-to-Nose | 55'-65' | 19'-34' | 10'-25' |
| Length, Wing-to-Tail | 45'-78' | 20'-40' | 9'-20' |
| Position of Engines | On-Wing | On-Wing/Fuselage | Concealed |
| Tailspan | 49'-92' | 23'-44' | 23'-28' |
| Tail Sweep, Leading | 116°-128° | 118°-122° | 94°-99° |
| Tail Sweep, Trailing | 97°-104° | 99°-104° | 85°-89° |
| Height, Overall | 38'-65' | 18'-34' | 9'-15' |

## References


[1] R. Bergevin and M. D. Levine. Generic object recognition: Building and matching coarse descriptions from line drawings. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 15(1):19–36, 1993.

[2] B. Bhanu and C.-C. Ho. Building hierarchical vision model of objects with multiple representations. In *Proc. SPIE on Applications of Artificial Intell. X: Machine Vision and Robotics, vol. 1708*, pages 663–674, Orlando, FL, April 1992.

[3] I. Biederman. Human image understanding: Recent research and a theory. *Comp. Vis. Grap. Image Proc.*, 32(1):29–73, 1985.

[4] R. A. Brooks. Symbolic reasoning among 3-dimensional models and 2-dimensional images. *Artificial Intell.*, 17:285–349, 1981.

[5] D. F. DeMenthon and L. S. Davis. Model-based object pose in 25 lines of code. In *Proc. DARPA Image Understanding Workshop*, pages 753–761, San Diego, CA, Jan. 1992.

[6] J. J. Koenderink. *Solid Shape.* Cambridge, MA: The MIT Press, 1990.

[7] B. Tversky and K. Hemenway. Objects, parts, and categories. *J. Expt. Psych.: General*, 113(2):169–193, 1984.

[8] A. J. Vayda and A. C. Kak. A robot vision system for recognition of generic shaped objects. *Comp. Vis., Graph., Image Understanding*, 54(1):1–46, 1991.

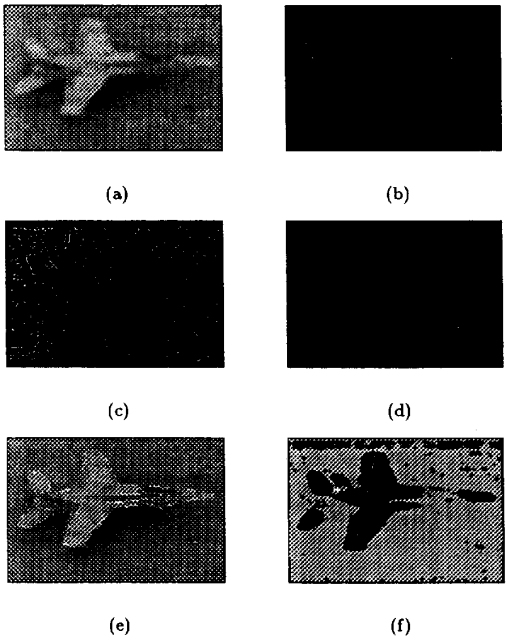



Figure 8: Results of low-level processing of the bottom ROI in Figure 7(c): (a) Original ROI image ($162 \times 240$). Extraction of thinned edges using different thresholds for edge magnitude: (b) $t = 225$, (c) $t = 50$. Results of feature extraction: (d) Detection of most salient edge contours, (e) Fitting straight lines to the contours of (d), (f) Segmented regions and extracted dominant axes for the largest foreground region.

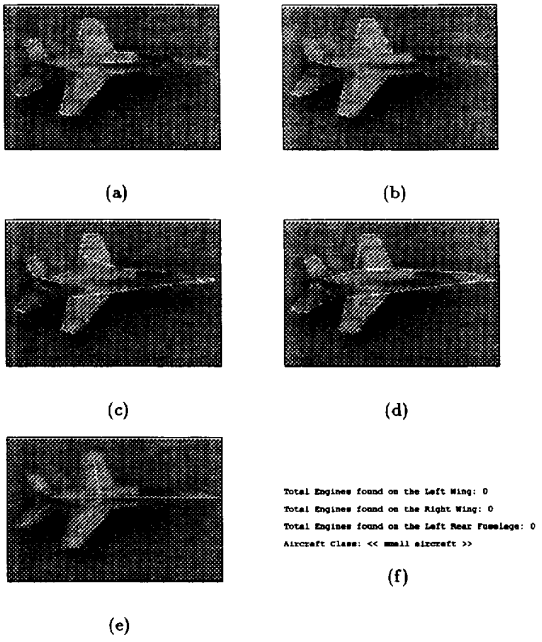



Figure 9: Results of qualitative object recognition: (a) Six convex groups of lines identified in Figure 8(e), (b) Trapezoid-like shapes identified using these groups, (c) Structural parts found during generic object recognition, (d) Refined structural parts that are also labeled, (e) Finding the skeleton of the shape, (f) Class recognition.

Table 2: Line segments for computing approximate pose. The pixel values are obtained from Figure 9(e) and the model dimensions from Table 1.

| Segment | Offset relative to $M_0$ | | Model dimension |
|---|---|---|---|
| | in pixels | in pixels | in meters |
| $i$ | $r_0 - r_i$ | $c_0 - c_i$ | |
| $M_0M_1$ | 0 | -130 | 3.5-11.0 |
| $M_0M_2$ | 50 | 8 | 5.0-11.8 |
| $M_0M_3$ | -57 | 33 | 5.0-11.8 |
| $M_0M_4$ | 0 | 45 | 2.7-6.0 |

Table 3: Approximate dimensions of the recognized subclass (small) of a generic aircraft corresponding to a set of best (lowest) G scores. The highlighted entry is the closest to the groundtruth, i.e., an F-18.

| Dimensional parameters | | | Goodness Score |
|---|---|---|---|
| $a$ (in meters) | $b$ (in meters) | $\gamma$ (in degrees) | $G$ |
| 8 | 6 | 95 | 0.59 |
| 11 | 6 | 115 | 0.59 |
| 3.5 | 5 | 93 | 0.49 |
| 6.5 | 9 | 93 | 0.53 |
| 9.5 | 12 | 93 | 0.53 |
| 8.5 | 6 | 115 | 0.59 |
| 10.5 | 6 | 115 | 0.59 |