



ELSEVIER

Image and Vision Computing 14 (1996) 323–338



Generic object recognition using multiple representations

Subhodev Das^a, Bir Bhanu^a, Chih-Cheng Ho^b

^aCollege of Engineering, University of California, Riverside, CA 92521-0425, USA

^bDepartment of Computer Science, University of Utah, Salt Lake City, UT 84112, USA

Received 6 September 1993; revised 22 September 1995

Abstract

Real-world image understanding tasks often involve complex object models which are not adequately represented by a single representational scheme for the various recognition scenarios encountered in practice. 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 derivation of hierarchical CAD models having multiple representations – concave/convex edges and straight homogeneous generalized cylinder – and their use 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.

Keywords: CAD-based vision; Feature extraction; Hierarchical representation; Model approximation; Object recognition

1. Introduction

The success of a 3D 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 indexing and search techniques to match the model and the data. In real-world image understanding (IU) problems like photointerpretation, there are additional factors which complicate the overall model-based object recognition process, such as occlusion, shadow, cloud cover, haze, seasonal variations, clutter, etc. Typically, CAD models of objects are used in IU tasks involving man-made objects. By and large, these CAD-based IU systems use a single representation scheme for the models, and a matching technique based on that representation for object recognition. However, for many applications, it is unlikely that a single representation-based recognition strategy would suffice for a variety of complex 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. Besides, the imaging conditions and the viewpoint location can affect the sensory data in such a way that the recognition of objects using fixed, detailed CAD models may be computationally

expensive. The use of generic or variable object models, each of which represent a set of distinct objects, can reduce the complexity of the recognition process.

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. The use of generic shape models in outdoor image interpretation has been limited to simple shapes like rectangles [1,2]. In comparison, our approach derives equivalent-shape models or approximations of complex CAD models for given resolution of the scene as the generic models for recognition. Hierarchical or parts/sub-parts structures based on the decomposition of a model's surface and/or volume are then represented using multiple schemes. Psychological studies [3,4] have also provided evidence of multiple representations of parts in human visual recognition. Apparently, such representations are used to guide the extraction of image features and, during the matching phase, admit cooperative matching strategies and simultaneous verification of parts of a hypothesized object. In addition to parts decomposition of individual models, the current approach assumes a hierarchical structure of the model database which consists of qualitative-to-quantitative descriptions of models. Generic class models (e.g. aircraft) are at the top of the hierarchy and are described

qualitatively, while fixed CAD models (e.g. Boeing 747) located at the terminal nodes have precise quantitative information; an object subclass (e.g. jumbo aircraft) at an intermediate level in the hierarchy has partial qualitative and variable quantitative specifications. The hierarchical database is assumed to have been obtained through a process of generalization of the fixed CAD models.

The paper is organized in the following way. Section 2 presents the related background and motivation behind the work reported in the paper. Derivation of equivalent-shape models from B-spline based CAD models and representation of the decomposed parts of these models using multiple schemes are described in Section 3. An algorithm for generic object recognition using these hierarchical models with multiple representations is presented in Section 4. Section 5 gives the details of implementation and the experimental results using real-world data and object class and subclass models. Section 6 presents the concluding remarks.

2. Background and motivation

Generic object recognition or classification of objects is accomplished using shape models of variable dimensions. In comparison, the traditional model-based object recognition or identification of objects is based on geometric models of fixed dimensions. Interestingly, categorization or the process of treating nonidentical stimuli as equivalent is performed in humans in a highly deterministic way and very often occurs at a basic level [5]. Basic categories are those which carry most distinguishing features, such as aircraft, bridge, house, tree, and typically are the ones which are first discriminated in an environment. Besides basic categories, psychological studies have also emphasized the importance of parts in recognizing visual objects. In this section, we will review the past work related to generic and parts-based object recognition. Following this, we will discuss our motivation for parametrically modifying CAD models depending on the viewing scale to obtain generic shape models and using multiple representations for these models and their components.

2.1. Background

One of the earliest generic object recognition systems is the ACRONYM [6]. It uses parametrized objects which represent families of objects characterized by sets of free parameters. Because of this representational scheme, all the features those are characteristics of a specialization are also present in the corresponding generic description. In ACRONYM, the reasoning about object classes and their specializations is carried out at the same time and failure to recognize the generalized class causes failure of

the subclass recognition. In spite of its representational elegance, the ACRONYM system is inadequate in addressing the real-world concerns posed by IU problems. It has no mechanism for automatically refining object models, and shadows, clutter and other nonidealities are not adequately handled within its framework. More recently, the PARVO system [7] has been introduced for fast and generic recognition of unexpected 3D objects. The approach is based on the Recognition By Components (RBC) theory [3] and generates a coarse parts-based volumetric description from a line drawing depicting a single view of an object. Unlike the ACRONYM, which uses quantitative parametrized models, the PARVO system utilizes coarse, qualitative models representing object classes. The INGEN system [8] has been used to interpret scenes containing objects for which only simple shape models such as parallelepipeds and cylinders are available but the sizes of the objects are unknown. The interpretation process is based on the available range data. Kadono et al. [9] have used hierarchical CAD models with size-variable planar surface patches to match multiple instances of a class of objects in multisensory images.

In CAD-driven vision, hierarchical or parts-based object descriptions are obtained by suitably decomposing the CAD models [10]. According to the psychological theory of Recognition-By-Components proposed by Biederman [3], complex objects are recognized not by the concepts of the whole object but by the objects' components described in terms of simple volumetric primitives. If some of these basic primitives, called geons, can be identified in their specified arrangement, then the identification of the object will be fast and accurate. In the studies by Tversky and Hemenway [4], parts are the basic level of human concepts not only because smaller parts are easier to deal with, but also because different parts are to be handled differently. They propose that part configuration underlies the various empirical operations of perception, behavior, and communication which converge at the basic level.

2.2. Motivation

Since the number of object classes is considerably less than that of specific objects in any domain, the complexity of object classification process using generic models is proportionately lower than that of object identification using specific models. Additionally, the results of the classification process can be used to constrain the identification search process, such as approximating the pose of specific objects. Generic CAD models which are obtained through a process of generalization of the common features of a class of specific CAD models are often too detailed for the kind of features which may be reliably extracted from images. Instead, the resolution limit imposed by the viewpoint location

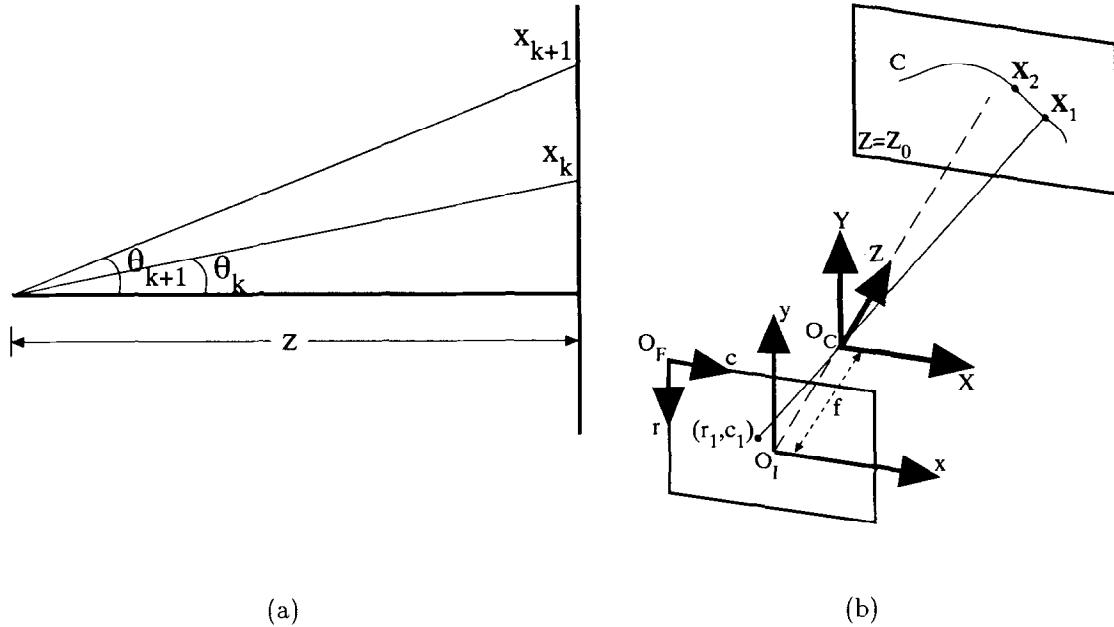


Fig. 1. Understanding view-dependent model representation. (a) Geometry illustrating lattice points x_k 's as function of a free variable z in one dimension; (b) perspective projection of a planar curve.

can be used to obtain simplified generic models. Under real-world conditions such as partial occlusion, the recognition of a complex object model is difficult unless the given model is structured into components. Each component is then detected in the input data and the evidence of the object is sought in the spatial relationships among the detected components.

To understand the need for viewpoint-dependent model representation, first consider the following 1D problem. Let x_0, x_1, \dots, x_N denote $N + 1$ lattice points in $[a, b]$ spaced Δx apart and let $\mathbf{f} = [f_0, f_1, \dots, f_N]^T$ be the observed values of a function $f(x)$ at these locations. Suppose $p(x) \in \mathcal{P}_N$ is a polynomial fit to the f_i 's: $p(x) = \sum_{j=0}^N a_j x^j$, where \mathcal{P}_N is the set of all polynomials of degree at most equal to N . The $N + 1$ coefficients $\mathbf{a} = [a_0, a_1, \dots, a_N]^T$ are obtained as unique solutions of the system of equations $\mathbf{a} = V^{-1} \mathbf{f}$, where V is the $(N + 1) \times (N + 1)$ Vandermonde matrix of the basis functions $1, x, x^2, \dots, x^N$ evaluated at the lattice points. Now, consider that the lattice point spacing, Δx , is a function of a free variable z as shown in Fig. 1(a). According to the figure, the relation between Δx and z is expressed as

$$\begin{aligned} \Delta x &= |x_{k+1} - x_k| = |\tan \theta_{k+1} - \tan \theta_k| \\ &\approx z |\theta_{k+1} - \theta_k| = z |\Delta \theta| \end{aligned} \quad (1)$$

under the conditions of $z \gg x$. In the context of model-based object recognition, the polynomial correspond to shape descriptors such as the components (axis, cross section, etc.) of a generalized cylinder (GC), while the variable z corresponds to the distance of the sensor from an instance of a model. The parameter $\Delta\theta$ is

determined by the resolution of the sensor and, therefore, it is fixed. From Eq. (1), it follows that the lattice spacing Δx increases as z increases subject to the condition that $\Delta\theta$ is constant. An increase in Δx for a fixed interval $[a, b]$ implies that there are fewer elements of the vector \mathbf{f} . Hence, it may be concluded that the number of coefficients of $p(x)$ gets smaller or the polynomial descriptor becomes smoother (of lower order) as z increases.

Generic object models accommodate the concept of adequate model description which can be supported by the sensory data. Space (3D) curves used to define these generic models correspond to the polynomial descriptor in the above 1D example, while z represents the viewing scale. Thus, utilizing the knowledge of the latter, it is possible to select appropriate (order of) space curves to define generic models for a given resolution. The notion of shape, as in shape recognition problems such as generic object recognition, is meaningful only if it is accompanied by a specified resolution at which the shape is to be assessed. This is well motivated by the fact that shapes appearing dissimilar at a high resolution can be identical at a low resolution [11]. We therefore emphasize the use of appropriately detailed models, to be called the diffused models, for the purpose of model-image matching. In these models, the degree of detail is determined by the viewing conditions or the imaging parameters. Nasr and Bhanu [12] and Bobick and Bolles [13] have advocated multiple, concurrent object descriptions for such applications as robot navigation where the descriptions are to be refined over time due to changes in the scene resolution.

Recognition of generic object models is further facilitated by organizing the individual models in a hierarchical fashion, a fact supported by past psychological

studies. A CAD-based vision system can acquire hierarchical models directly from a CAD database. Each subpart of a decomposed CAD model can then be independently described using more than one representation. The use of multiple representations facilitates cooperative interaction among different matching strategies, which may be pursued to exploit the benefit of the individual representational schemes. Since the goal of parts-based recognition is to verify, in the image data, the spatial arrangement of the parts, the multiple representation scheme leads to a more efficient recognition, particularly in unconstrained environments, than a single representation-based recognition method. As each representation imposes its own set of constraints on the image observables, the different representations only enrich the overall constraint set which increases the possibility of detecting the parts reliably and reduces the number of alternate hypotheses those need to be pursued during verification of the spatial arrangement of the detected parts.

When the image understanding task requires identification of specific objects, the object recognition process must continue beyond classification in terms of generic models. In such situations, access to more specific models is necessary. This requirement is satisfied by arranging the model database in a hierarchical fashion (in addition to hierarchical decomposition of individual models). The generic object models are coordinated to be near the top of the arrangement, while more specific models are disposed further down the hierarchy. Using this database, reasoning about objects and their classes can be cascaded without requiring the presence of the same features (for matching) at all levels. The matching process can also search a lower level for distinguishing features should a categorization be not possible at a particular level because of the lack of suitable features. Thus, the flow of control during matching is bi-directional – between a generalized class and its more specialized subclasses.

3. Building generic object models with multiple representations

We begin with the derivation of the conditions, as functions of the imaging geometry, under which view-dependent or diffused generic model descriptions may be obtained for given CAD models. This is followed by a description of creating multiple representation-based hierarchical components of diffused models.

3.1. Conditions for diffused model representation

Consider the perspective projection of a planar curve as illustrated in Fig. 1(b). Let the planar curve, C , be described by the implicit equation $p(X, Y) = 0$, where

$p(\cdot)$ is a multivariate polynomial such that $p(\cdot) \in \mathcal{P}_N$. Also, 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 . The corresponding frame-based coordinates (i.e. in O_{Frc} system) are

$$\mathbf{r}_1^{(h)} = {}^F\mathbf{A}_C \mathbf{X}_1^{(h)} \quad \text{and} \quad \mathbf{r}_2^{(h)} = {}^F\mathbf{A}_C \mathbf{X}_2^{(h)} \quad (2)$$

where ${}^F\mathbf{A}_C$ is the homogeneous transformation matrix given by

$${}^F\mathbf{A}_C = \begin{bmatrix} 0 & fk_y & \bar{r} & 0 \\ -fk_r & 0 & \bar{c} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3)$$

Here, (\bar{r}, \bar{c}) denote the frame coordinates of O_I , f is the focal length of the imaging system, and k_x and k_y are internal parameters of the sensor. Now, referring to Fig. 1(b),

$$\Delta r = |\mathbf{r}_2 - \mathbf{r}_1| = \frac{fk_y}{Z_0} |X_2 - X_1| \quad \text{and} \\ \Delta c = |c_2 - c_1| = \frac{fk_x}{Z_0} |Y_2 - Y_1|. \quad (4)$$

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 $Z = Z_0$ must be

$$\Delta X = |X_2 - X_1| \geq \frac{Z_0}{fk_y} \quad \text{and} \quad \Delta Y = |Y_2 - Y_1| \geq \frac{Z_0}{fk_x}. \quad (5)$$

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 ‘ ϵ -neighborhood’ ([11], based on point set topology) for our shape metric. By definition, ϵ -neighborhood of a point is a spherical region of radius ϵ centered at that point and the ‘spherical ϵ -neighborhood’ of an arbitrary region is the union of all ϵ -neighborhoods of points belonging to that region. Thus, the ‘distance’ between two shapes is the smallest value of ϵ 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 the distance between them is within the limit of the specified resolution.*

In our case, the smallest ϵ corresponding to the distance between C and its diffused model is $\epsilon = [(\Delta X)^2 + (\Delta Y)^2]^{1/2}$. Let the diffused model of $p(\cdot)$ be represented by the implicit equation $q(X, Y) = 0$, $q(\cdot) \in \mathcal{P}_M$ and $M < N$. The Vandermonde matrix corresponding to

the original model is

$$V = \begin{bmatrix} 1 & X_0 & Y_0 & \cdots & X_0^m & \cdots & X_0^i Y_0^j & \cdots & Y_0^m & \cdots & X_0^n & \cdots & Y_0^n \\ \vdots & \ddots & \ddots & \cdots & \ddots \\ 1 & X_N & Y_N & \cdots & X_N^m & \cdots & X_N^i Y_N^j & \cdots & Y_N^m & \cdots & X_N^n & \cdots & Y_N^n \end{bmatrix}$$

where $(X_0, Y_0, Z_0), \dots, (X_N, Y_N, Z_0)$ are any $N+1$ points belonging to the curve $p(\cdot)$ and $0 \leq i+j \leq m$. The rank, $r(V)$, of V which is an $(N+1) \times (N+1)$ matrix with $N = \binom{n+2}{2} - 1$, is thus $r(V) \leq N$. Let these $N+1$ points be perturbed such that the resultant Vandermonde matrix is

$$\tilde{V} = \begin{bmatrix} 1 & X_0 - \epsilon_0 & Y_0 - \nu_0 & \cdots & (X_0 - \epsilon_0)^m & \cdots & (Y_0 - \nu_0)^m & \cdots & (Y_0 - \nu_0)^n \\ \vdots & \ddots & \ddots & \cdots & \ddots & \cdots & \ddots & \cdots & \ddots \\ 1 & X_N - \epsilon_N & Y_N - \nu_N & \cdots & (X_N - \epsilon_N)^m & \cdots & (Y_N - \nu_N)^m & \cdots & (Y_N - \nu_N)^n \end{bmatrix}$$

where ϵ and ν are the perturbation amounts such that $|\epsilon_i| < \Delta X$ and $|\nu_i| < \Delta Y$, $i = 0, \dots, N$. The Vandermonde matrix, \tilde{V} , corresponding to the diffused model is an $(M+1) \times (M+1)$ submatrix of \tilde{V} containing elements of the form $X^i Y^j$, $0 \leq i+j \leq m$, such that $M = \binom{m+2}{2} - 1$ and $r(\tilde{V}) \leq M$. Thus, the necessary condition for the existence of an m -th order diffused model of $p(\cdot)$ is that $|\det(\tilde{V})| = 0$ and according to Hadamard's inequality this condition is satisfied iff \tilde{V} has a zero column vector. If ΔX (ΔY) is such that $|X - \epsilon| \ll 1$ ($|Y - \nu| \ll 1$) and $(X - \epsilon)^m \rightarrow 0$ ($(Y - \nu)^m \rightarrow 0$), then $|\det(\tilde{V})| \rightarrow 0$. Hence, it follows that coarser the resolution, i.e. ΔX and ΔY large, better is the chance of identifying ϵ and ν such that the m -th order diffused model exists. The sufficient conditions for the existence of an m -th order diffused model, $q(\cdot)$, of $p(\cdot)$ are $r(\tilde{V}) \leq M$ and $\max \|p(X, Y) - q(X, Y)\| < \epsilon$. These conditions can be used in the derivations of diffused models as discussed next.

3.2. Multiple representations of diffused models

The generic CAD models used in this work are created using the Alpha_1 solid modeling system [14], developed at the University of Utah. Alpha_1 models the geometry of solid objects by representing their boundaries using NURBS (NonUniform Rational B-spline Surfaces). Using Alpha_1, generic objects are designed with various geometric operators, such as extrude, bend, stretch, warp, etc., or combinations of them using boolean operations, set union, difference, intersection, etc. [10]. Fig. 2(a) shows an airplane CAD model whose decomposition is shown in Fig. 2(d). The decomposition is obtained by first deriving a polyhedral approximation of the B-spline model (Fig. 2(b)) of Fig. 2(a), and then finding the concave edges (Fig. 2(c)) which 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 [10]. Several other techniques for representing surfaces by edges are given in Bhanu et al. [15]. Once the CAD models are derived, appropriate diffusion techniques can then be applied to each object model.

Given a 3D curve and a diffusion limit ϵ based on resolution, it is possible to obtain a (nonunique) diffused 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. Suppose the 3D curve has the following parametric representation which is the most common representation for curves and surfaces in CAD: $\mathbf{x}(t) = [x(t) y(t) z(t)]^\top$, $t \in [a, b] \subset \mathcal{R}$, where $\dot{\mathbf{x}}(t) = d\mathbf{x}/dt \neq 0$. Let the curve be reparametrized (without affecting its shape) using the arc length parameter s , where $s = s(t) = \int_a^t \|\dot{\mathbf{x}}\| dt$ and $\|\cdot\|$ denotes the norm of a vector. Also, let $\mathbf{t}-\mathbf{n}-\mathbf{b}$ be the Frenet frame at a point s along the curve; \mathbf{t} is the tangent vector to the curve at s , i.e. $\mathbf{t} = \mathbf{x}'$, \mathbf{n} is the principal normal vector, i.e. the unit vector normal to the contour and lying in the plane containing \mathbf{t} and \mathbf{t}' , and \mathbf{b} is the binormal vector, i.e. the unit vector normal to \mathbf{t} and \mathbf{n} at s . Here, ' $'$ ' denotes differentiation with respect to s . The Frenet–Serret formulas relating these vectors to their derivatives are given by [16]

$$\mathbf{t}' = \kappa \mathbf{n}, \quad \mathbf{n}' = -\kappa \mathbf{t} + \tau \mathbf{b}, \quad \mathbf{b}' = -\tau \mathbf{n}, \quad (6)$$

where the terms κ and τ are called the curvature and torsion, respectively. The curvature may be defined both in terms of the original parameter t and in terms of s as [17]

$$\kappa = \kappa(t) \left[\frac{\|\dot{\mathbf{x}} \times \ddot{\mathbf{x}}\|}{\|\dot{\mathbf{x}}\|^3} \right], \quad \text{and} \quad \kappa = \kappa(s) = \|\mathbf{x}''\|. \quad (7)$$

Similarly, the torsion is defined as [17]

$$\tau = \tau(t) = \frac{\det[\dot{\mathbf{x}}, \ddot{\mathbf{x}}, \ddot{\mathbf{x}}]}{\|\dot{\mathbf{x}} \times \ddot{\mathbf{x}}\|^2} \quad \text{and} \quad \tau = \tau(s) = \frac{1}{\kappa^2} \det[\mathbf{x}, \mathbf{x}'', \mathbf{x}''']. \quad (8)$$

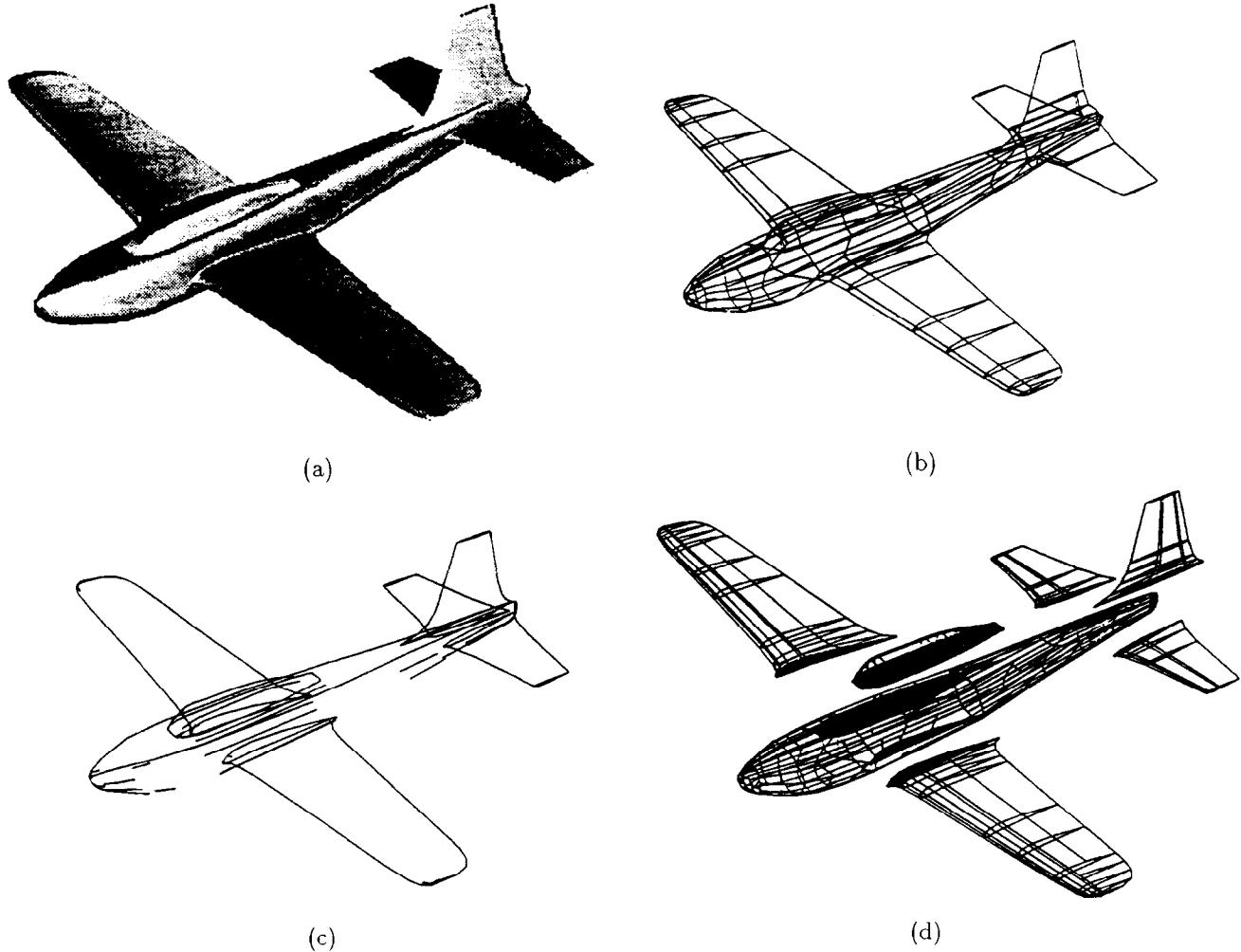


Fig. 2. A generic airplane model. (a) B-spline model of an airplane, (b) polyhedral approximation of (a), (c) edge detection using (b), (d) decomposition of (b) along concave edges of (c).

The above 3D curve is subjected to an infinitesimal deformation such that the new (deformed) curve is represented as $\mathbf{x} + \delta\mathbf{x}$, where $\delta\mathbf{x} = \alpha t + \beta\mathbf{n} + \gamma\mathbf{b}$. If λ denotes the amount of deformation along the curve, i.e. $\delta(ds) = \lambda ds$, then the curvature of the new curve is obtained as

$$\kappa + \delta\kappa = \mu' + \kappa(1 - \lambda) + \tau\nu. \quad (9)$$

Here, $\lambda = \alpha' - \kappa\beta$, $\mu = \beta' + \alpha\kappa - \gamma\tau$, and $\nu = \gamma + \beta\tau$, while “’” denotes differentiation with respect to s . Thus, given the original and the diffused curves in parametric form and hence their respective curvatures and torsions according to Eqs. (7) and (8), an iterative deformation of the original curve in steps of $\delta\mathbf{x}$ allows one to determine the smallest degree of the diffused curve in terms of the degree of the original curve using Eq. (9). During this iterative process, it must be ensured that the total deformation ρ of the original curve, where $\rho = \sum_i \|\delta\mathbf{x}_i\|$, i being the number of iterations, is such that $\rho < \epsilon$ for a given value of ϵ , the allowed perturbation amount.

To illustrate the above diffusion procedure, consider a GC-based description of a CAD model whose axes and

cross-sections are typically described by low-order polynomials. If an approximate estimate of the viewing scale is known (this is often available for applications like photointerpretation), then the scene resolution is obtained using Eq. (5), which in turn is used to determine the tolerance, ϵ , of diffusion. 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. Fig. 3(a) shows the cross sections of GCs. Fig. 3(b) shows the axes of GC representations of the subparts of the original (non-diffused) airplane model and Fig. 3(c) describes the axes of the corresponding subparts of the diffused model using the SHGC representation. The model of Fig. 3(c) can be further approximated depending on ϵ -value to obtain a diffused model in which small SHGCs, i.e. ones with axes

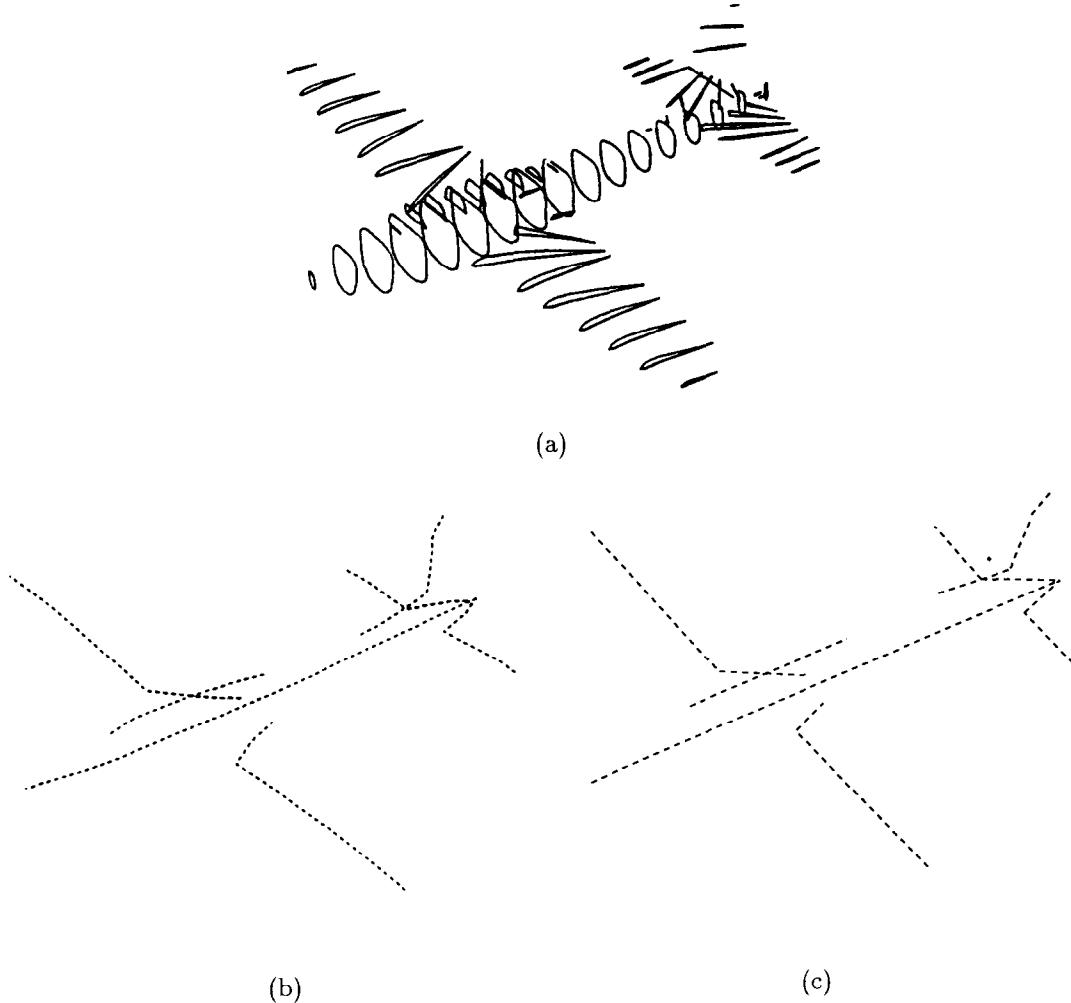


Fig. 3. Generalized cylinder representation of subparts of an airplane. (a) Cross-sections of generalized cylinders, (b) axes of generalized cylinders, (c) axes of straight homogeneous generalized cylinder-approximations of generalized cylinders in (b).

lengths nearly equal to or smaller than ϵ , are ignored, e.g. the SHGCs connecting the wings or the tails to the fuselage in Fig. 3(c).

Multiple-representation descriptions are maintained for each of the decomposed subparts using polyhedral approximation, concave/convex edges, curvature extrema, surface normals, and generalized cylinders. Unlike Constructive Solid Geometry or other techniques commonly applied for representing object decompositions in CAD systems, the selected representational schemes are quite suitable for computer vision applications. Most man-made objects encountered in typical IU applications are conveniently represented using one or more of these selected methods. For example, GCs are well suited for objects having axes of symmetry; similar objects have similar axes and cross sections, while any small difference between similar objects will be reflected by a small difference in the axes or cross sections. Polyhedral representation is widely used because of its simplicity and good support of both geometrical and topological information. Two examples of multiple

representations of object models are shown for the airplane model: the edge points of Fig. 2(c), and the generalized cylinder representation of Fig. 3.

4. 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 for model feature extraction from images. Spatial arrangement of the subparts extracted from images as indicated in the object model specification is verified through accumulation of evidence during matching. The recognition process initiates with the most generic, diffused CAD models. Guided by the hierarchical object model database, 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

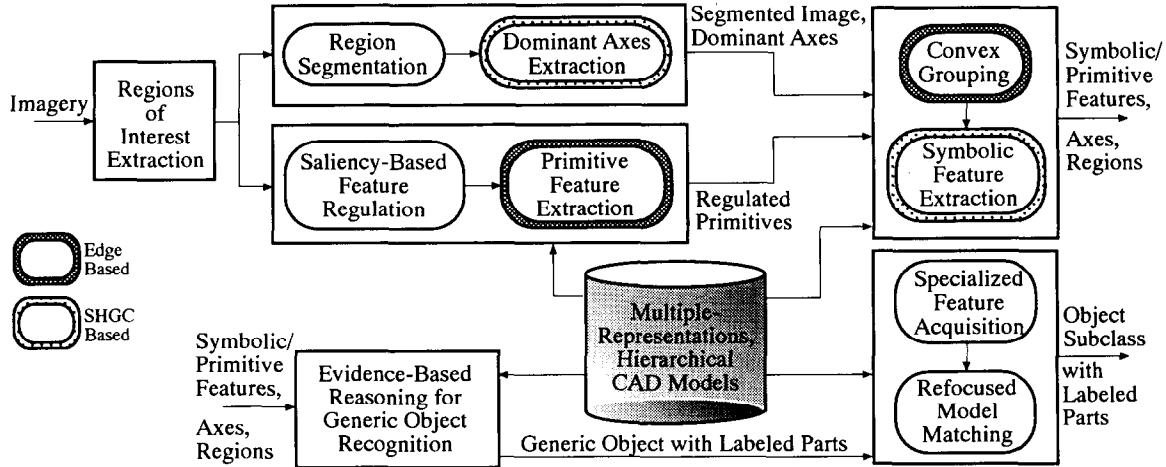


Fig. 4. A flow diagram of the generic object recognition algorithm.

generic object recognition consists of the steps shown in Fig. 4 and the overall process is supervised by a controller.

The input to the recognition system is a 2D 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) which 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.

4.1. SHGC representation-based dominant axes extraction

Axes are important features for recognizing objects represented by generalized cylinders. Of these, 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 [18,19]. 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 2D shapes of objects are used to extract the dominant axes from the images. Such shapes are identified by segmenting image regions which 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 [20].

To extract dominant axes of an SHGC-based representation, we note that the curvature at an end point of the axis of an SHGC is inversely proportional to the distance from its origin of scaling. The approach for 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. Consequently, an axis can be generated by connecting the midpoints of the opposite ends of an

SHGC contour at the points of high curvature. 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 which are not the centers of adjacent clusters. Lines whose significant portions are not contained within the segmented region are ignored.

4.2. Edge representation-based primitive feature extraction

Detection of intensity edges is the initial step for most feature extraction processes. Intensity edge pixels are identified by applying multiple thresholds to the output of an edge detector and are thinned to one-pixel width. Next, long edge segments which 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 Fig. 4). The details of the linked-segment extraction step appear in [21].

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, which is similar to the one proposed in [22], is a set of regulated intensity-edge segments. Now, the model edges intersect in 3D corners whose projections are 2D corners. Thus, in addition to lines, our algorithm also detects corners by obtaining gradient and curvature measurements at pixels in the grey scale image [23].

4.3. 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. These symbolic features correspond to the different subparts of a generic object.

4.3.1. 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 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.

4.3.2. 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 Fig. 2(a), subparts can be represented using LRSHGCs for certain degree of diffusion as in aerial image interpretation.

The following are some useful properties of RSHGC and LSHGC representations which are utilized during the symbolic feature extraction:

Property 1 *The occluding contour of an RSHGC is always planar in an end or side view. In an oblique view, the occluding contour of a Linear RSHGC always lies in a plane [24].*

Property 2 *The 2D image of the occluding contour of a Linear RSHGC is a line.*

Property 3 *The 2D image of a solid corner formed by two contour generators of a Linear RSHGC or by two intersecting Linear RSHGCs is a corner.*

Based on these properties, the contour of an LRSHGC can be represented using a convex group of lines. Consequently, it is possible to extract the contours of LRSHGCs from convex groups of lines. In our aircraft example, convex groups of lines are used to extract subparts like wings, tails, and the rudder sections of the generic aircraft model.

4.4. 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 those 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 those can be found in the input data. The final output are the identified symbolic parts of the generic object.

4.5. Refocused matching

The labeled symbolic parts are now used to direct the image-based search for more localized features which are available at lower levels of the database hierarchy. The symbolic features may have associated qualitative or

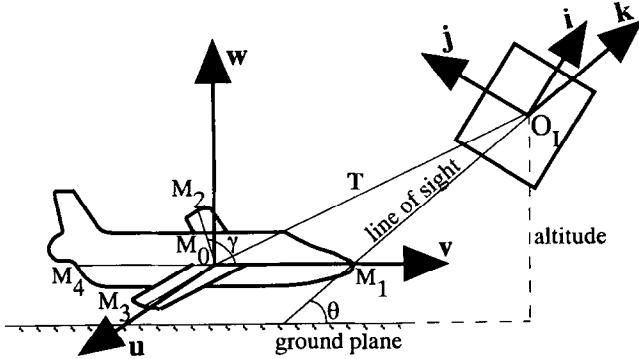


Fig. 5. The coordinate systems relating an object model to its projection for pose computation. The vectors in the systems \mathbf{u} - \mathbf{v} - \mathbf{w} and \mathbf{i} - \mathbf{j} - \mathbf{k} are unit vectors.

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 [25].

Referring to Fig. 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

$$\begin{aligned} r_0 - r_i &= \frac{fk_y}{\bar{Z}}(Y_0 - Y_i) - s_i \mathbf{i} \cdot \mathbf{M}_0 \mathbf{M}_i, \\ c_0 - c_i &= \frac{fk_x}{\bar{Z}}(X_0 - X_i) = s_x \mathbf{j} \cdot \mathbf{M}_0 \mathbf{M}_i, \end{aligned} \quad (10)$$

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 3×3 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. $i = 1, \dots, m$, the corresponding systems of equations are represented using matrix vector notation as

$$\mathbf{r} = A\mathbf{I}, \quad \mathbf{c} = AJ. \quad (11)$$

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} and \mathbf{k} . 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 $\mathbf{O}_1 \mathbf{M}_0$. The translation vectors corresponding to all other model points can be similarly obtained. These, together with T , allow updating the estimates of $\mathbf{M}_0 \mathbf{M}_i$'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 available initially. These improved estimates can subsequently be used in the identification step.

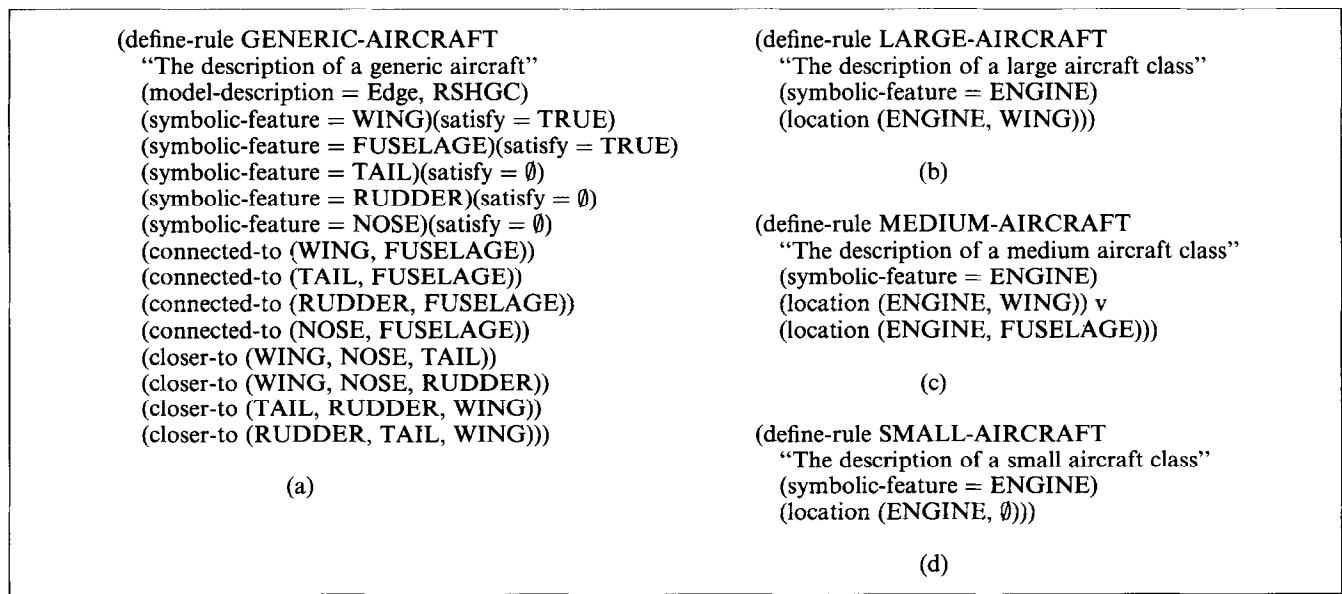
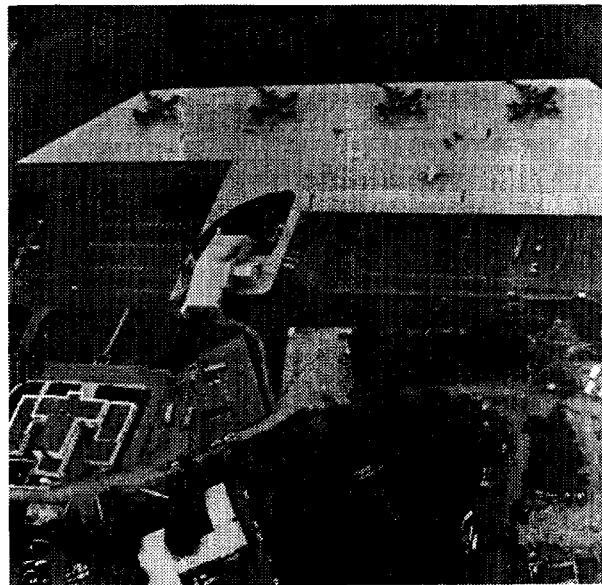


Fig. 6. Descriptions of hierarchical CAD models with multiple representations. (a) A generic aircraft, and the three aircraft classes (b) large, (c) medium (d) small.

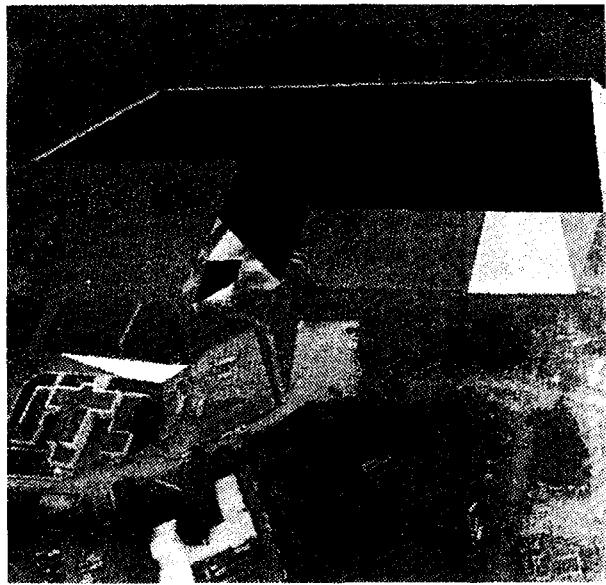
5. Experimental results

The results reported in this paper are based on one generic object – aircraft – and its three subclasses – large, medium, and small (see Fig. 6). Fig. 7(a) shows an aerial photograph ($4K \times 4K$) which has several aircraft – four C-130's and one F-18. Using the multi-resolution focusing approach, several ROIs are identified as shown in Figs. 7(b)–(d). These ROIs are analyzed by the object recognition system in succession. Here, we

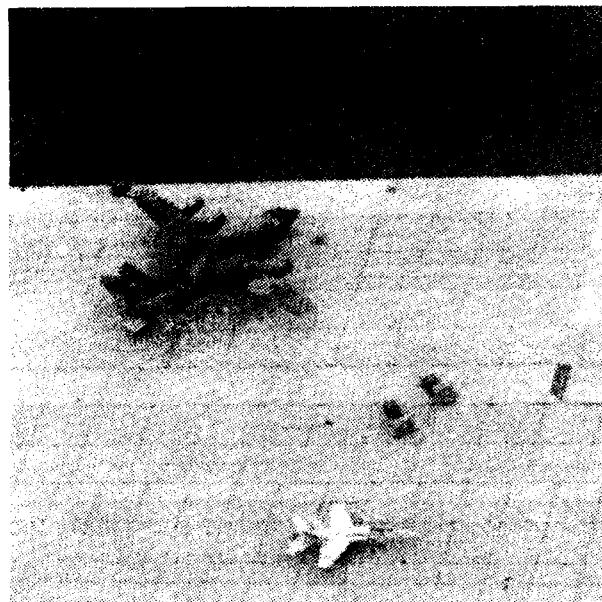
present the results of analyzing the bottom ROI (162×240) from Fig. 7(d) (contains the F-18 aircraft). The ROI and the output of the multi-threshold edge detection step are shown in Figs. 8(a)–(c). In our implementation, we have selected five threshold (t) values which are fixed for all images. The result of extracting globally salient edge contours is presented in Fig. 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



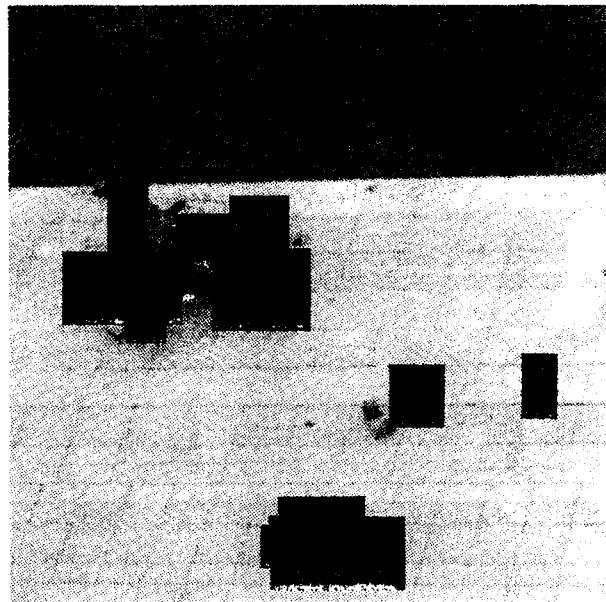
(a)



(b)



(c)



(d)

Fig. 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).

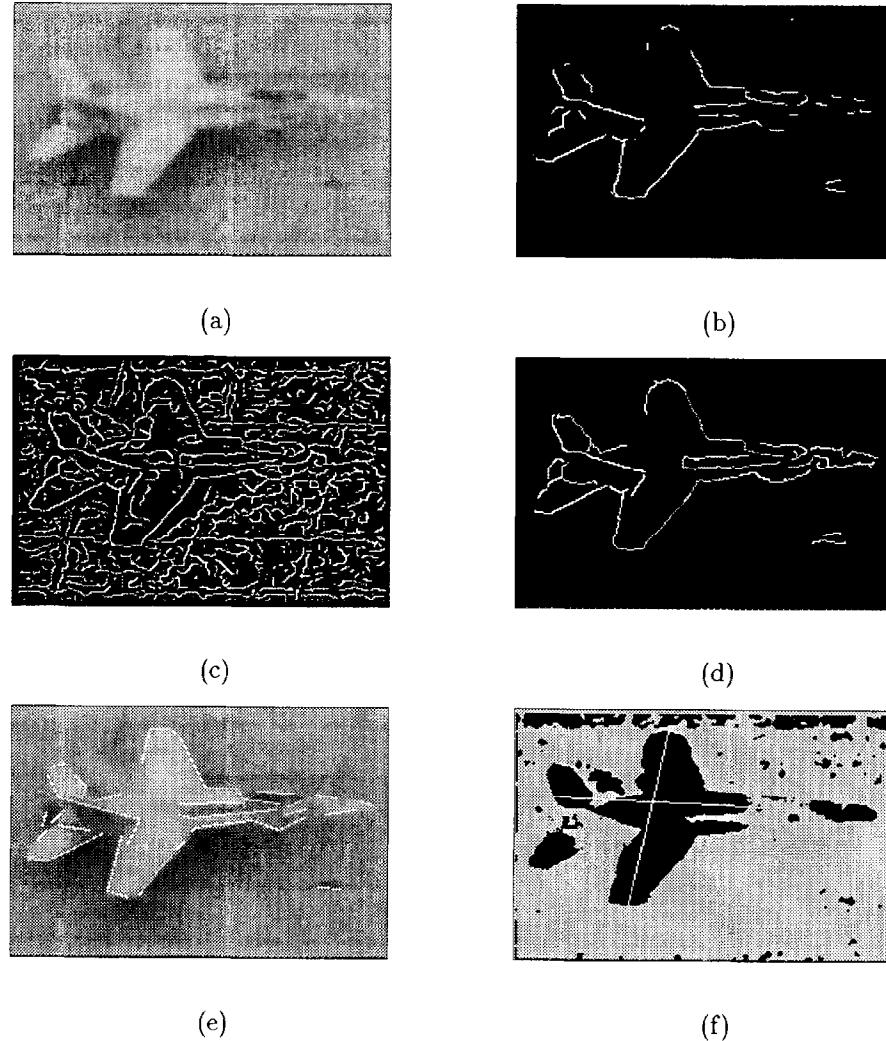


Fig. 8. Results of low-level processing the bottom ROI in Fig. 7(d). (a) Original ROI image (162×240). Extraction of thinned edges using different thresholds for edge magnitude: (b) $t = 225$, (c) $t = 50$. Results of feature extraction: (d) detection of most salient edge contours, (e) fitting straight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

result of line fitting to the salient structures is shown in Fig. 8(e). Segmented regions and the dominant axes of regions are shown in Fig. 8(f). Fig. 9(a) shows the six convex sets of lines identified using convex grouping procedure. These are used to extract trapezoid-like features shown in Fig. 9(b) which are the symbolic descriptions of some of the subparts, like 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 as specified in Fig. 6 have been satisfied, can their ensemble be recognized as a generic aircraft. The identified subparts are shown in Fig. 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 Figs. 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 (Fig. 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 (Fig. 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 Fig. 5. The range of Z (refer to Section 4.5) measured along the LOS is obtained as 798–801 m based on Table 1 data and

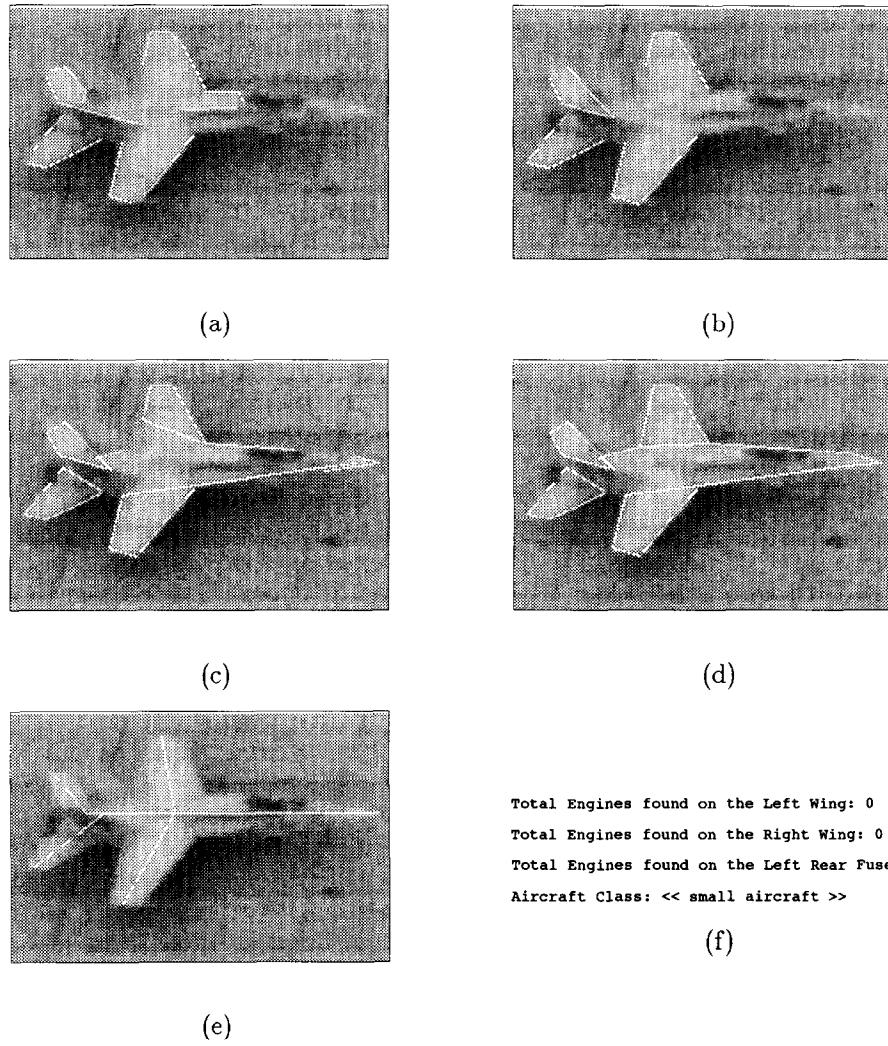


Fig. 9. Results of qualitative object recognition. (a) Six convex groups of lines identified in Fig. 8(e), (b) trapezoid-like shapes identified using these groups, (c) structural parts found during generic object recognition, (d) refined structural parts that are also labeled, (e) finding the skeleton of the shape, (f) class recognition.

Table 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'

Fig. 5. The four line segments, indicated as M_0M_1 , M_0M_2 , M_0M_3 and M_0M_4 in Fig. 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. (11) consisting of the coordinates of the model points M_1 , M_2 , M_3 , M_4

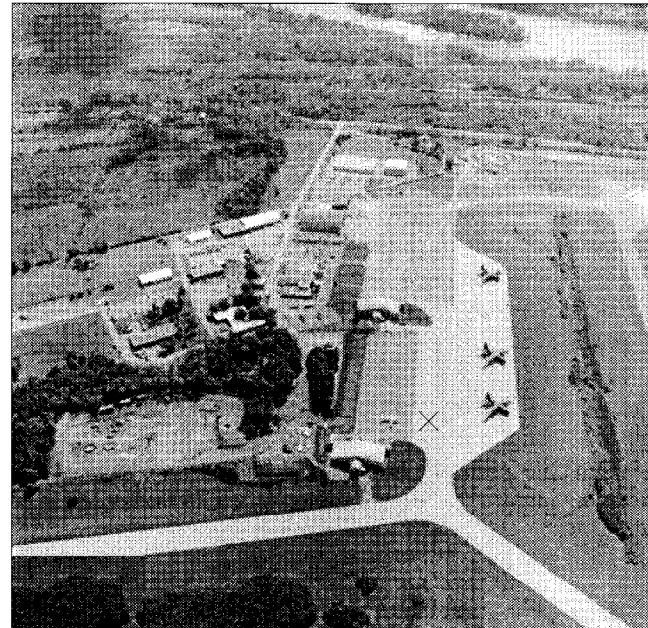
Table 2
Line segments for computing approximate pose. The pixel values are obtained from Fig. 9(e) and the model dimensions from Table 1

Segment	Offset relative to M_0		Model dimension in meters
	in pixels	in pixels	
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

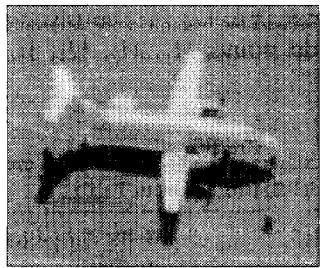
Table 3

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

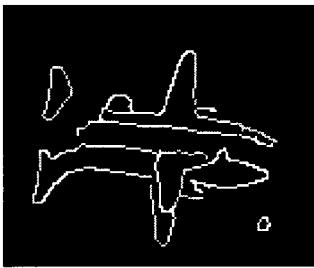
Dimensional parameters			Goodness Score G
a (in meters)	b (in meters)	γ (in degrees)	
8	6	95	0.59
11	6	115	0.59
3.5	5	93	0.49
6.5	9	93	0.53
9.5	12	93	0.53
8.5	6	115	0.59
10.5	6	115	0.59



(a)



(b)



(c)

Fig. 10. Aerial view of an airfield. (a) Original image ($4K \times 4K$), (b) a ROI image (120×140) of the aircraft marked with a \times in (a), (c) most salient contour configuration.

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 $93^\circ \leq \gamma \leq 115^\circ$. 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 3×3 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 , γ 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 γ values corresponding to the interval of G around this mode agrees with the ground-truth, i.e. the dimensions of an F-18 aircraft. The ranges of a , b and γ 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 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.

A second aerial image is shown in Fig. 10(a). In this example, we present the results of generic recognition at the subclass level. The results involve analysis of the ROI of Fig. 10(b). The top-level (most salient) configuration of perceptually salient contours is shown in Fig. 10(c). The linear segments of this structure are displayed in Fig. 11(a). The ROI is segmented into three sets – shadow, object and background – using recursive application of the segmentation algorithm [20]. The dominant axes of the largest object region are shown in Fig. 11(b). The shadow lines in Fig. 11(a) are identified using bimodality test of neighborhood histograms and are removed. Next, convex groups of the non-shadow lines are used to extract trapezoid-like symbolic features. The details of the non-shadow symbolic feature extraction process appear in [21]. Structural parts identified during evidence-based recognition are shown in Fig. 11(c) and are further refined as shown in Fig. 11(d). Finally, the class of the generic aircraft is determined to be large based on engine locations as seen in Fig. 11(e).

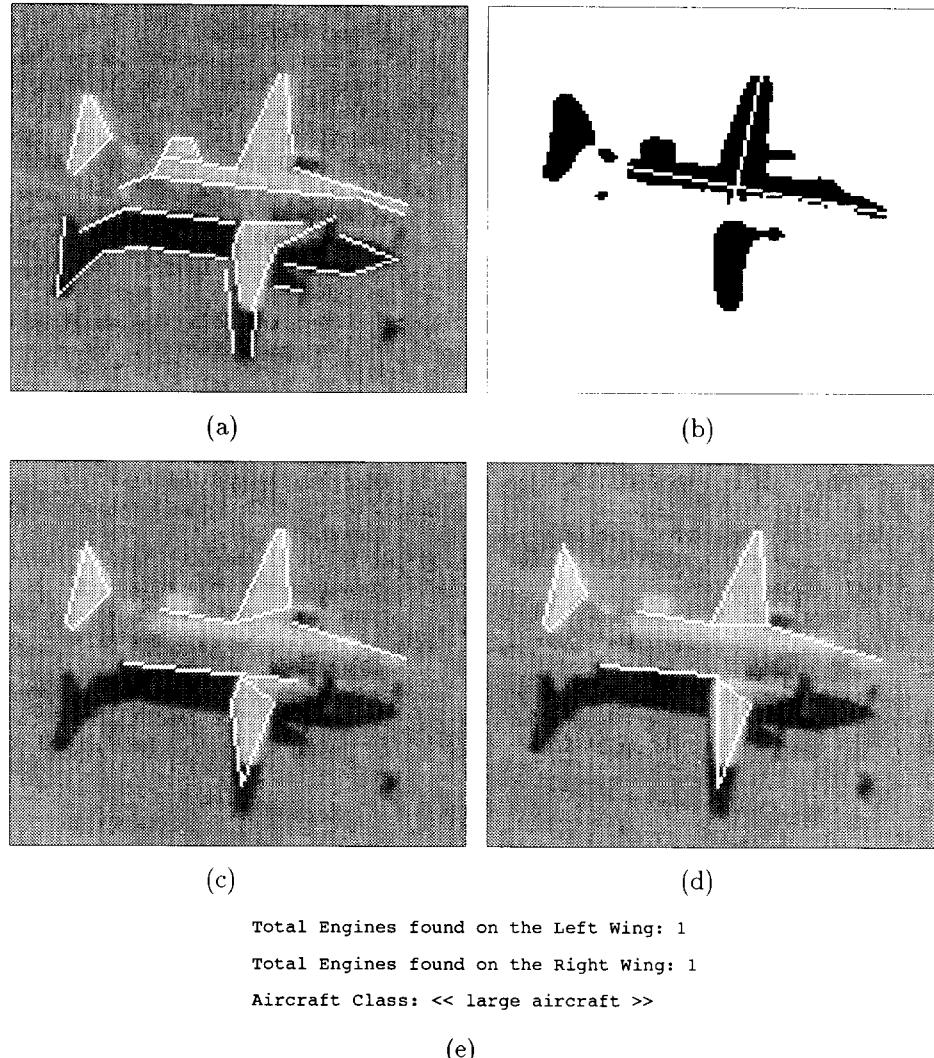


Fig. 11. Results of feature extraction. (a) Straight lines fitted to the most salient contours, (b) dominant axes for the largest object region. Results of qualitative object recognition: (c) structural parts found during generic object recognition, (d) refined structural parts, (e) class recognition.

6. Conclusions

In this paper, we have presented an approach to recognition of generic 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 as specified by the generic model. The novel aspects of our work are parametrically modifying CAD models depending on the viewing scale to obtain generic shape models, and the use of multiple representations for the hierarchical subparts of these models. The multiple representations have been demonstrated to be useful for extraction of object features and as constraints during recognition. The generic object recognition strategy emphasized in this paper can serve as the initial step for any model-based object recognition technique. In particular, the results obtained here can be used to perform increasingly focused search of the

precise models in the database. We have demonstrated this by seeking further classification of a recognized generic object and computing the approximate dimensions of the object.

References

- [1] P. Fua and A.J. Hanson, Using generic geometric models for intelligent shape extraction, *Proc. DARPA Image Understanding Workshop*, 1987, pp. 227–233.
- [2] A. Huertas and R. Nevatia, Detecting buildings in aerial images, *Computer Vision, Graphics and Image Processing*, 41 (1988) 131–152.
- [3] I. Biederman, Human image understanding: Recent research and a theory, *Comp. Vis. Grap. Image Proc.*, 32(1) (1985) 29–73.
- [4] B. Tversky and K. Hemenway, Objects, parts, and categories, *J. Expt. Psych.: General*, 113(2) (1984) 169–193.
- [5] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson and P. Boyes-Braem, Basic objects in natural categories, *Cognit. Psychol.*, 8(3) (1976) 382–439.

- [6] R.A. Brooks, Symbolic reasoning among 3-dimensional models and 2-dimensional images, *Artificial Intell.*, 17 (1981) 285–349.
- [7] 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) (1993) 19–36.
- [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) (1991) 1–46.
- [9] K. Kadono, M. Asada and Y. Shirai, Context-constrained matching of hierarchical CAD-based models for outdoor scene interpretation, *Proc. IEEE Workshop on Directions in Automated CAD-Based Vision*, Maui, HI, June 1991, pp. 186–195.
- [10] B. Bhanu and C.-C. Ho, Building hierarchical vision model of objects with multiple representations, *Proc. SPIE on Applications of Artificial Intell. X: Machine Vision and Robotics*, 1708, Orlando, FL, April 1992, pp. 663–674.
- [11] J.J. Koenderink, *Solid Shape*, MIT Press, 1990.
- [12] H. Nasr and B. Bhanu, Landmark recognition for autonomous mobile robots, *Proc. IEEE Int. Conf. on Robotics and Automation*, Philadelphia, PA, April 1988, pp. 1218–1223.
- [13] A.F. Bobick and R.C. Bolles, The representation space paradigm of concurrent evolving object descriptions, *IEEE Trans. Pattern Anal. Machine Intell.*, 14(2) (1992) 146–156.
- [14] Alpha_1 Research Group, Department of Computer Science, University of Utah, Salt Lake City, Utah. *Alpha_1 Users Manual*, 1986.
- [15] B. Bhanu, S. Lee, C.-C. Ho and T.C. Henderson, Range data processing: Representation of surfaces by edges, *Proc. 8th Int. Conf. Pattern Recognition*, Paris, France, October 1986, pp. 27–31.
- [16] M.P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.
- [17] G. Farin, *Curves and Surfaces for Computer Aided Geometric Design – A Practical Guide*, Academic Press, 1988.
- [18] J. Ponce, D. Chelberg and W.B. Mann, Invariant properties of straight homogeneous generalized cylinders and their contours, *IEEE Trans. Pattern Anal. Machine Intell.*, 11(9) (1989) 951–966.
- [19] H. Sato and T.O. Binford, *BUILDER-I: A system for the extraction of SHGC objects in an edge image*, Proc. DARPA Image Understanding Workshop, San Diego, CA, January 1992, pp. 779–791.
- [20] B. Bhanu and R.D. Holben, Model-based segmentation of FLIR images, *IEEE Trans. Aerospace Elec. Sys.*, 26(1) (1990) 2–11.
- [21] S. Das, B. Bhanu, X. Wu and R.N. Braithwaite, A system for aircraft recognition in perspective aerial images, *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL; December 1994, pp. 168–175.
- [22] D.G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer, 1985.
- [23] H.-H. Nagel, Displacement vectors derived from second-order intensity variations in image sequences, *Computer Vision, Graphics, and Image Processing*, 9 (1983) 203–214.
- [24] S.A. Shafer, *Shadows and Silhouettes in Computer Vision*, Kluwer, 1985.
- [25] D.F. DeMenthon and L.S. Davis, Model-based object pose in 25 lines of code, *Proc. DARPA Image Understanding Workshop*, San Diego, CA, January 1992, pp. 753–761.