



PII: S0031-3203(97)00061-7

A SYSTEM FOR MODEL-BASED OBJECT RECOGNITION IN PERSPECTIVE AERIAL IMAGES

SUBHODEV DAS* and BIR BHANU

College of Engineering, University of California, Riverside, CA 92521-0425, U.S.A.

(Received 9 December 1993; in revised form 12 June 1997)

Abstract—Recognition of objects in complex, perspective aerial imagery is difficult because of occlusion, shadow, clutter and various forms of image degradation. This paper presents a system for aircraft recognition under real-world conditions. The particular approach is based on the use of a hierarchical database of object models and involves three key processes: (a) The *qualitative object recognition* process performs heterogeneous model-based symbolic feature extraction and generic object recognition; (b) The *refocused matching and evaluation* process refines the extracted features for more specific classification with input from (a); and (c) The *primitive feature* extraction process regulates the extracted features based on their saliency and interacts with (a) and (b). Experimental results showing the qualitative recognition of aircraft in perspective, aerial images are presented. © 1998 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved

Object recognition Qualitative vision CAD-based vision Photointerpretation
 Perceptual grouping Salient feature extraction Shadow analysis

1. INTRODUCTION

Photointerpretation (PI) or extraction of intelligence from image data, particularly aerial imagery, has been an important application domain of image understanding (IU) techniques for about two decades. The primary goal of PI is image exploitation to aid reconnaissance tasks, such as airfield, port, and troop movement monitoring. An IU system for PI is typically required to identify buildings, aircraft, ships, ground vehicles, bridges, and storage facilities, using geometric (e.g. CAD-based) models. Of these, the first two object classes — buildings and aircraft — have received the maximum attention in the literature. Recognition of buildings has proven to be relatively easier than that of aircraft because of the structural simplicity of the former. Although there have been several aircraft recognition systems proposed in the past, very few of these have actually addressed the concerns of real-world, such as occlusion, shadow, cloud cover, haze, seasonal variations, clutter, and various other forms of image degradation. To illustrate the difficulty of the problem, we show typical aircraft images in Fig. 1 where recognition has to be accomplished under varying contrast and in presence of shadows.

However, there exists no viable approach to date that can work satisfactorily on all these example images.

In this paper, we describe an end-to-end IU system for aircraft recognition which has been demonstrated to be effective in presence of shadows, clutter, and low image contrast. Our system uses a hierarchical representation of aircraft models consisting of generic aircraft, aircraft classes (e.g. jumbo aircraft), specific aircraft (e.g. Boeing 747), and aspects of a specific aircraft. Such representation is in terms of qualitative-to-quantitative descriptions that vary from advance concepts (e.g. aircraft wing) to primitive geometric entities (e.g. points, lines) and allow increasingly focused search of the precise models in the database. To account for image variabilities, our system exploits heterogeneous models, such as those of image segmentation, camera/platform, sun, shadow to derive symbolic features in a robust manner. Finally, the system regulates the extracted primitive features based on their saliency, an ability that helps to distinguish object model features from image clutter.

Section 2 describes the background and motivation behind the work reported in this paper. Section 3 describes the novel features of our aircraft recognition system. Section 4 presents the details of an algorithm that integrates feature refinement and object classification. Section 5 gives the specifics of implementation and the experimental results for qualitative recognition of aircraft using real-world data. Section 6 presents concluding remarks.

*Current address: PVI, 47 Hulfish St, Princeton, NJ 08542, U.S.A.

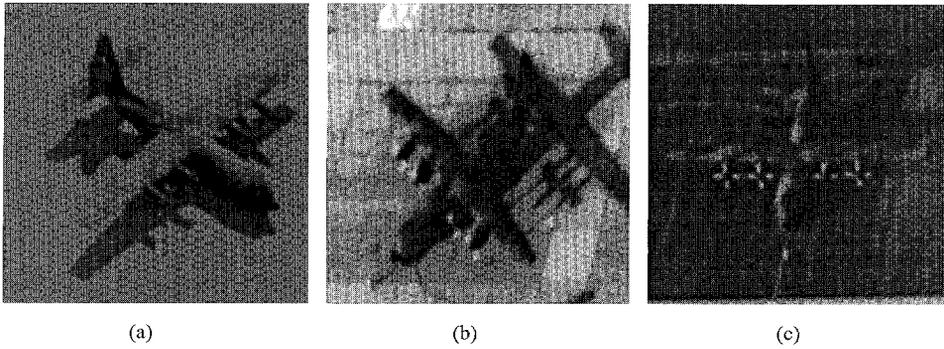


Fig. 1. Representative aircraft recognition scenarios: (a) shadow and medium contrast, (b) shadow and low-to-medium contrast, and (c) shadow and very low contrast.

2. BACKGROUND AND MOTIVATION

There is a variety of object recognition and classification techniques which can possibly be applied to the domain of aircraft recognition. However, the algorithms reviewed in this section are those which have been applied to the specific problem of aircraft identification in both 2-D and 3-D. Following this, we discuss the motivation behind our particular approach to aircraft recognition.

2.1. Background

The different approaches to aircraft recognition which have been proposed so far can be broadly classified into the following categories:

- *Moment invariant techniques*—These techniques use moment invariant features of the aircraft silhouette and silhouette border to perform the classification task. The advantage of these techniques is that these invariants are unaffected by rotational, translational, and scaling differences between an object model and its observed image. The disadvantage is the sensitivity of the invariants to mass distribution inside the silhouette, occlusion, clutter, noise, and other image abnormalities.
- *Syntactic/semantic grammar techniques*—These approaches use linguistic pattern recognition techniques to analyze shapes and classify aircraft using piecewise linear border approximations. Using a set of terminal symbols to represent image primitives (lines or arcs) and relationships (parallel, collinear, right angles) between these primitives, a grammar is derived to specify allowable combinations of these primitives to construct complete aircraft borders. Other information, e.g. mean, variance, etc., about a particular object is represented by the semantics introduced into the grammar. Thus, recognition of individual aircraft reduces to the task of parsing a set of words (image primitives) to create legal sentences (aircraft borders). The advantages of syntactic/semantic grammar over mo-

ment invariants are that the former allows specification of local structure rather than global shape and the variations in object shapes can be explicitly incorporated into the models, thereby reducing the sensitivity to noise. However, these approaches are also inadequate in handling occlusion and clutter.

- *Fourier descriptor techniques*—In these approaches, the shape of the aircraft's closed contour in the image plane is represented using a Fourier descriptor (FD) which is subsequently used to recognize future instances of the aircraft. The principle of FD is that the boundary of a closed planar figure can be expressed as a function of some variable and repeating this process multiple times will produce a periodic function that can be expressed in a Fourier series. This series is the Fourier descriptor of the planar figure. Various normalization procedures are used to derive feature sets which are invariant with respect to starting point, rotation, translation, and scale. The advantage of FDs is that partial shape matching in presence of occlusion can also lead to complete classification. However, there may be instances when the normalization for deriving invariant features is not uniquely determined. Other factors affecting FD-based approaches are number of sampling points used on the contour, uniformity of sample spacing, number of FDs used, quantization error, and the amount of perturbation of the contour.
- *Model and knowledge-based techniques*—These techniques seek to represent an aircraft using advance concepts in a hierarchical part-subpart fashion, where the lowest-level representation is usually in terms of image primitives. The recognition process begins by locating these image primitives and then by combining them in a forward- or backward-chaining fashion using the system's model and knowledge base. The advantage of these techniques is that they rely on spatially local features which can be extracted from the sensory data with relative ease. Such features can lead to relatively robust model matching in presence of

noise, occlusion, and missing data, when compared to the global shape representations, by incorporating appropriate object, sensory, and contextual information. The main drawbacks of these techniques are the “knowledge acquisition bottleneck” and the real-time implementation.

- *Other techniques*—These are the techniques which do not exactly belong to any one of the categories described above.

Summaries of the specific algorithms which are representative of these groups appear in Table 1.

The “knowledge-free” techniques are inadequate for real scenes as they treat the object of interest in isolation from the rest of the image. Also, these approaches have very limited capability to handle clutter, shadow, occlusion, etc. Among the knowledge-based techniques, the ACRONYM system by Brooks⁽¹⁾ is closest to the system described in this paper. However, the primitive feature extraction process in ACRONYM is independent of the subsequent analysis. So errors in preprocessing of primitives persist throughout analysis. Also, ACRONYM does not possess any hypothesis verification step and the complexity of its constraint-based modeling is limited. All of these factors limit its application to images which are relatively simple photometrically, e.g. free of shadows and clutter.

2.2. Motivation

Most model-based object recognition approaches rely on some form of segmentation of the raw image primitives into meaningful groups. However, good segmentation cannot be guaranteed for real-world, complex images. This difficulty is compounded by the fact that a specific aircraft type can have very different appearances depending on the imaging conditions, e.g. the hercules aircraft in Figs 1(b) and (c). Consequently, the recognition of aircraft using 2-D or 3-D quantitative object models is very difficult and almost none of the proposed methods have been shown to be effective for the diverse scenarios encountered in PI. Since one has little control on the imaging conditions and hence on the types of features which may be obtained, a more effective approach would be to adopt a multi-step recognition approach that proceeds from identifying generic object classes to specific object instances. A hierarchy of model database in which the terminal nodes are specific object models of geometric entities and successive higher levels are symbolic conceptualization of the lower-level nodes is well suited for this purpose.

One of the key factors of poor recognition performance in real-world scenarios is the presence of shadows as seen in Fig. 1. However, shadows provide valuable 3-D information and have been successfully utilized in recognizing buildings in aerial photographs,^(2,3) but have found little attention in the context of aircraft recognition. Shadows can be help-

ful in the identification of image features which are caused by raised structures, such as buildings, poles, or aircraft, and also in the determination of their 3-D orientations, such as a vertically oriented structure (e.g. a pole) as opposed to a horizontally oriented structure (e.g. an aircraft). Occlusion is another issue which a robust aircraft recognition system needs to address. In aerial imagery, occlusion is primarily due to self-occlusion or close parking of aircraft. The shadow and occlusion issues signify that aircraft recognition is essentially a 3-D problem.

Feature selection is critical for successful object recognition in practical scenarios. For aircraft recognition, this is related to distinguishing between image features which are caused by the aircraft structure and those which constitute the clutter. One way to determine the relevant features is to identify the perceptually salient image contours (linked intensity edge segments). As psychological evidence indicate, very often such contours are associated with objects which could be of interest or which attract our visual attention immediately. Detection of such salient contours in aerial images, which could very well be aircraft contours, should considerably simplify the segmentation and recognition tasks. Previous attempts at aircraft recognition have focused on a single feature type, either edges (lines, closed boundaries, etc.) or regions (blobs, silhouettes, etc.). Some IU systems have incorporated both, but dynamically selected only one at any stage of processing, e.g. COBIUS.⁽⁴⁾ However, since the extraction processes for these two types of features seek to maximize different criterion functions and are, therefore, affected by the image variabilities in different ways, the overall recognition process is benefited by combining these features.

Extraction of salient image primitives and subsequent formation of symbolic descriptions do not guarantee successful recognition in a single processing step. This is because a single decision threshold on the measure of saliency may allow too many or too few primitives for the symbolic feature extraction step, thereby, inciting too many possibilities in the recognition outcome or none at all! On the other hand, if the degree of saliency is determined by the “generic” properties of a class of objects, then it is possible to have multiple decision thresholds. With an iterative recognition scheme, a different threshold in each iteration will allow only the “generic” features associated with the object “class” of the iteration to be extracted. For example, “wing” symbolic features which are certain to distinguish an aircraft class from most of other object classes encountered in an aerial imagery should be extracted to recognize an instance of an aircraft using long, low-curvature, salient contours in the initial iteration of an aircraft recognition cycle. Thus, further improvement in recognition is possible by integrating feature (both primitive and symbolic) selection and image interpretation processes, such that the interpretation process is driven by the most salient contours first, followed by the less salient ones. Such

Table 1. Summary of various aircraft recognition techniques

Techniques	References	Descriptions
Moment invariant	Gupta and Srinath ⁽¹⁶⁾	2-D; vector of a sequence of moment invariant functions computed from contour
	Gupta and Srinath ⁽¹⁷⁾	2-D; moment function based on distances between each contour pixel and object centroid
	Dudani <i>et al.</i> ⁽¹⁸⁾	3-D; invariant features computed from silhouette and its border for different views of each aircraft; Bayes and distance-weighted <i>k</i> -NN classification
	Reeves <i>et al.</i> ⁽¹⁹⁾	3-D; normalized moment invariants that are less sensitive to noise than conventional invariants; performance comparable to FDs with/without noise
Syntactic/semantic grammar	Tang and Huang ⁽²⁰⁾	2-D; large, useful structures obtained by removing redundant terminal symbols; localization without classification
	Davis and Henderson ⁽²¹⁾	2-D; hierarchical approach using only grammatically correct fragments at all levels with constraint propagation; classification into general categories
Fourier descriptor	Lin and Chellappa ⁽²²⁾	2-D; estimation of FD for the complete contour from partial data
	Wallace and Wintz ⁽²³⁾	3D; normalized FDs for global shape descriptions
	Wallace <i>et al.</i> ⁽²⁴⁾	3-D; FDs for local shape descriptions
	Gorin ⁽²⁵⁾	3-D; combines individual classification results from multiple frames to refine accuracy over time
	Gorman <i>et al.</i> ⁽²⁶⁾	3-D; partial recognition of occluded or overlapping objects using FDs of local features
	Chen and Ho ⁽²⁷⁾	3-D; elliptic FDs that are less sensitive to contour perturbation than regular FDs; reduced set of near neighbors in NN classification for speed
Model and knowledge based	Ming and Bhanu ⁽²⁸⁾	2-D; Explanation-based learning for model acquisition and refinement; Conceptual clustering for classification
	Brooks ⁽¹⁾	3-D; specific and generic objects, partially specified scene and camera models; iteration of prediction, description, and interpretation from coarse object subpart and class interpretations to fine distinctions among subclasses and precise 3D quantification
	Moldovan and Wu ⁽²⁹⁾	3-D; hierarchical classification using object skeleton, boundary, surface, volume, and ancillary data; top-down reasoning from higher abstraction level of object details to lower level of greater object-related information
	COBIUS ⁽⁴⁾	3-D; hierarchical representation of objects and constraints; dynamic selection of region or edge segmentation for initial interpretation, followed by model-based resegmentation to extract expected objects
Other	Ma <i>et al.</i> ⁽³⁰⁾	2-D; normalized shape descriptors faster than FDs and requiring less storage
	Gibbon ⁽³¹⁾	2-D; weighted chord functions to represent angle chords of an object
	Ben-Arie and Meiri ⁽³²⁾	3-D; matching of <i>n</i> -ary relational graphs
	Thompson and Mundy ⁽³³⁾	3-D; vertex-pairs as invariant features
	Chien and Aggarwal ⁽³⁴⁾	3-D; quadtree and octree-based object representation; identification using occluding contours
	Reeves and Taylor ⁽³⁵⁾	3-D; recognition based on contour, silhouette, and range imagery

integration will also facilitate the usage of the interpretation results of a previous integration step to direct the feature selection and interpretation processes of a later step. For example, classification of a generic aircraft using “wing” features should make the search for “engine” features easier and should facilitate further classification based on the latter features.

3. SYSTEM DESCRIPTION

The system for model-based aircraft recognition is schematically shown in Fig. 2. It has four key elements: (a) a qualitative-to-quantitative hierarchical object model database, and three recognition sub-processes which utilize these models, (b) saliency-based regulation of low-level features to be used in the

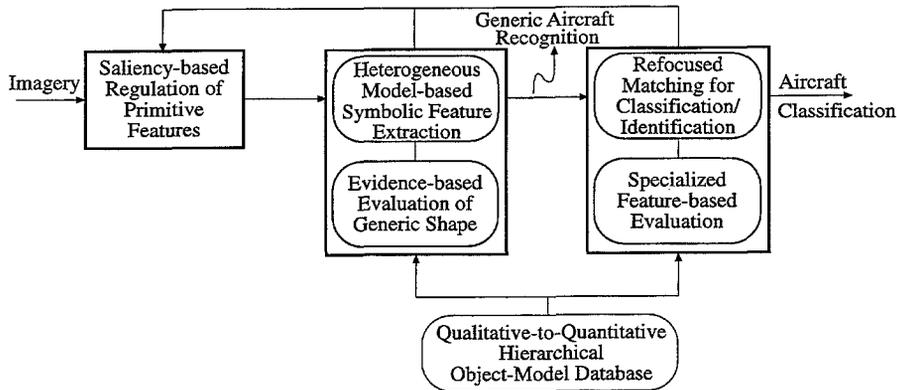


Fig. 2. A schematic of the model-based aircraft recognition system.

subsequent steps of recognition in an incremental fashion, (c) model-based symbolic feature extraction and evaluation that utilizes these regulated low-level features and heterogeneous models of image segmentation, shadow casting, and image acquisition, and (d) refocused matching for finer object classification. These features are now described in detail and their novelties are clearly identified.

3.1. Hierarchical object model database

There are two important considerations in designing a generic-to-specific hierarchical database from the point of view object recognition: the choice of features to represent a particular object class and the matching process.

Our system uses a representation of structural models in which the choice of features is driven by the discriminating power of these features among objects at the same level of the hierarchy. Additionally, the selected features (of an object model) are ranked according to their relative importance in recognizing that particular object and this order is followed during evidence accumulation. The top level of the hierarchy contains the shape (qualitative) attributes of a generic aircraft in terms of its structural subparts. The progressively deeper levels embody more specific knowledge that becomes completely quantitative at the terminal nodes, i.e. location of geometric models. A partial hierarchy illustrating this particular database structure is shown in Fig. 3.

Using the hierarchical 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 model features should a categorization be not possible at a particular level because of the lack of suitable features. For example, according to Fig. 3, the engine is not an essential structural feature to detect a generic aircraft and is, therefore, not a part of the generic aircraft model, but

it is required for subclass discrimination. However, this feature can be used for verifying a generic aircraft. Thus, the flow of control during matching is bi-directional — between a generalized class and its more specialized subclasses. The use of hierarchical object descriptions is proving to be useful for scene understanding tasks in general.⁽⁵⁾

Compared to our approach, ACRONYM⁽¹⁾ employs a representation scheme in which all the features which are characteristics of a specialization are also present in the corresponding generic description, except that these feature values are replaced by symbols (quantifiers) in the generic model. Matches to parts of an object class are automatically carried down to matches to the corresponding parts of a specialization of that object class. In other words, the control flow is always from the generalized to the specialized classes and failure to recognize the generalized class causes failure of the subclass recognition.

3.2. Saliency-based regulation of low-level features

Real-world, complex scenarios are characterized by large amount of image clutter which may mislead any system into false recognition. Clutter rejection is difficult using local image measures unless these measures have strong saliency, e.g. color. One way to distinguish an object from its background clutter is by its perceptual saliency.⁽⁶⁾ Humans can visually attend to salient structures almost immediately without scanning an entire image. The different saliency measures can be length, curvature, or contrast of primitive features.

Our approach utilizes simple local measures of saliency based on the strengths of detected edge pixels, and lengths and local curvatures of edge segments. As a first step, edge pixels at multiple thresholds are extracted. This is motivated by the fact that *no* single threshold is suitable for all the different images which may be encountered in practice, let alone for the different parts of a single image. Edge pixels are

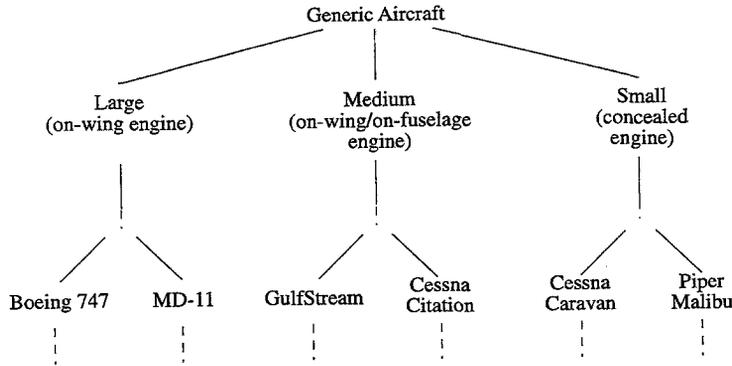


Fig. 3. A partial hierarchy of generic-to-specific aircraft.

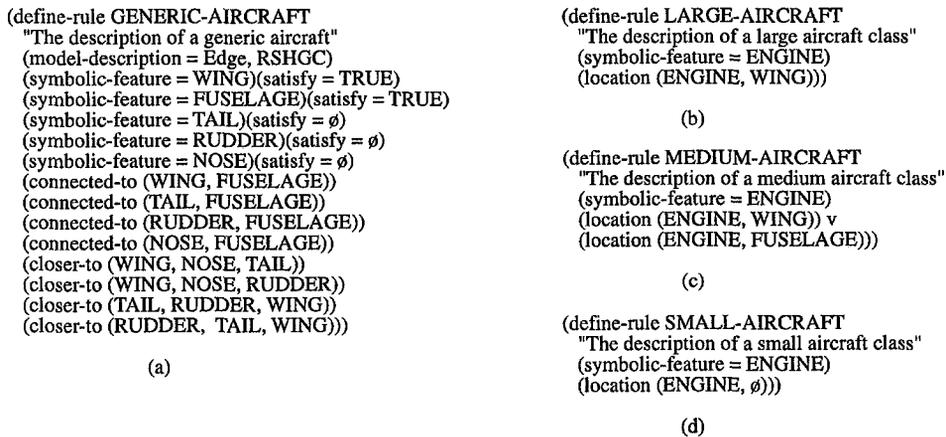


Fig. 4. Simplified examples of symbolic features: (a) A generic aircraft, and the three aircraft classes, (b) large, (c) medium, (d) small.

grouped based on the local curvature values and the continuity of edge segments. In addition, the flow of low-level features which are used to derive the symbolic features is regulated based on the saliency. Regulation may also be based on the "specialized" nature of the features as required by the refocused matching process.

The use of salient features is different from the locally focused features.⁽⁷⁾ The latter are suitable for industrial applications where images are far less complex and the models have certain unique characteristics.

3.3. Model-based symbolic feature extraction and evaluation

Perceptually grouped primitives which are typically employed in object recognition systems are based on local measures, such as proximity, collinearity, parallelism⁽⁸⁾ and can describe only simple shapes using primarily linear features. Symbolic features are derived from mappings of primitive features under

geometrical and physical constraints. However, by incorporating more domain-specific knowledge into these constraints, symbolic features can be tuned to the detection of specific object classes, such as aircraft.

Domain knowledge is embedded in the production rules describing the symbolic features. There are two types of production rules: qualitative, which measures the qualitative properties of symbolic features; quantitative, which computes values for the corresponding symbolic features. The production rule definitions of a generic aircraft and its three subclasses are illustrated in Fig. 4. To extract these features, the recognition search process is initiated at the top level of the database hierarchy and is looped through the production rules of the node being visited in a goal-decomposition fashion until a rule is encountered whose conditions require mapping of primitives.

The evaluation of the extracted features involves the verification of the global semantic shape components of the generic model. It utilizes heterogeneous models: edge/gray scale-based model of image segmentation, models of shadow casting process, and

models of image acquisition. Although, the use of multiple sources of information is routine for multi-sensor-based scene interpretation,^(5,9) such practice is uncommon for the single sensor-based approaches. Also used in our system are the dominant axes which characterize the shape of the generic aircraft class. During evaluation, feedback from the generic recognition module to the feature regulation module (see Fig. 2) helps to acquire additional low-level features in the event of recognition failure or low recognition confidence. Integrating feature acquisition and recognition is essential for robust object recognition. However, this aspect has *not* been addressed by previous model-based recognition techniques.

3.4. Refocused matching

This module is responsible for further classification of an object whose category has been determined by the generic recognition process. Functionally, refocused matching is an iterative recognition process attempting to match a reduced set of image data and a reduced set of object models, where reduction is with respect to the input image data and the entire object model library. In our approach, data reduction is achieved by deriving symbolic features which are more “focused” or localized with respect to a particular level of our database hierarchy. Usually, the symbolic features of the “generic” model identified in the previous iteration step guide the search for image primitives required for deriving these focused symbolic features during current iteration. In addition, the derived symbolic features may be subjected to mensuration.

The key difference between this step and multi-resolution approaches to recognition is that the former views the object model database at increasing

resolution instead of the image. It may be noted that in the absence of any higher-level control knowledge, the resampling process may exclude important object features while accommodating clutter.

4. ALGORITHM

The emphasis of this paper is on qualitative recognition of aircraft using the system described in the previous section. Initially, a solid modeling system is used to build B-spline CAD models of aircraft with *multiple representations*.⁽¹⁰⁾ In this work, we utilize two representations: polyhedral edge-based (approximations of B-spline CAD models) and straight homogeneous generalized cylinder (SHGC). Given an input 2-D image and ancillary data about the imaging parameters and scene conditions, the algorithm for qualitative recognition of aircraft consists of the steps illustrated in Fig. 5. First, the regions of interest (ROIs) likely to contain aircraft are identified in the high resolution input image by performing an intelligent search. The search begins with a low resolution version of the original image and identifies different image parts which have a large concentration of perceptual groups of primitive feature types characteristic of the targeted generic class. Next, each of these subimages are examined at increasingly higher resolution for the same purpose until the original resolution is attained. Further details of this step appear in.⁽¹¹⁾

The qualitative recognition of aircraft within each identified ROI is accomplished by two major algorithmic processes: heterogeneous model-based feature extraction, and integration of salient feature extraction and evidence-based model matching. The former process obtains symbolic features of the generic aircraft and its classes using region- and edge-based

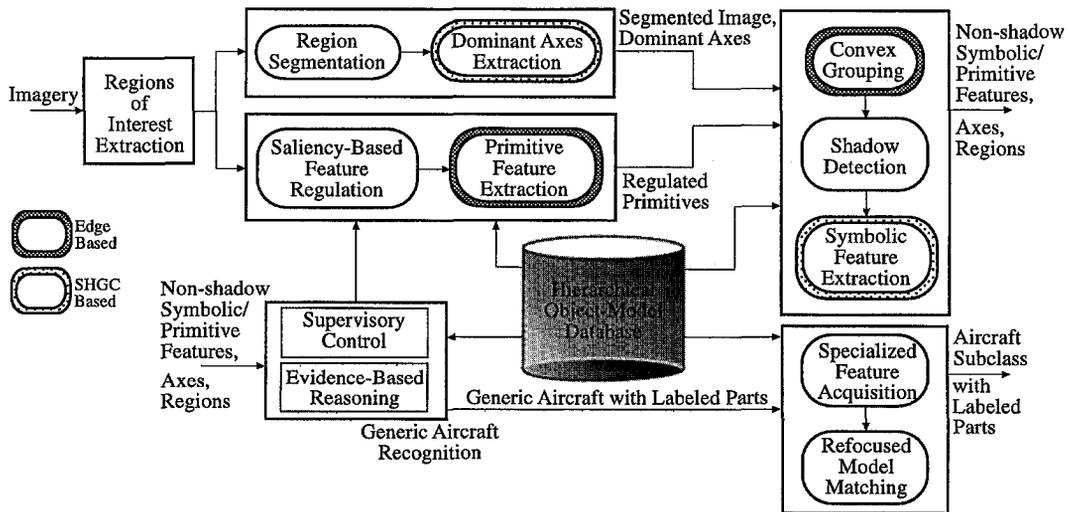


Fig. 5. A flow diagram of the qualitative aircraft recognition algorithm.

primitive features. It involves heterogeneous models of image segmentation utilizing edge/gray scale information, of shadow casting process, and of image acquisition. The integration process provides a feedback path from the model matching module to salient edge feature regulation module (see Fig. 5). The feedback control is used to acquire additional primitive/symbolic features in the event of recognition failure or low recognition confidence. In the following, we describe the individual steps of these two processes in further detail.

4.1. Salient edge feature regulation

Edge pixels are detected in an input ROI by applying multiple thresholds. This is motivated by the fact that *no* single threshold is suitable for all the different images which may be encountered in practice. We illustrate the effect of different thresholds on the output of an edge detector in Fig. 6. Clearly, a trial-and-error approach can only find a satisfactory threshold, but it is not guaranteed for all situations. Instead, we emphasize the extraction of perceptually salient contours from *any* given set of edge pixels obtained using multiple thresholds.

Our approach to identifying perceptually salient contours is based on finding long, smooth edge segments which are made up of high-magnitude edge pixels. This is formulated as a problem of finding an edge segment of length N starting at a terminal pixel, corresponding to $s = 0$ (s being the segment parameter), and subject to the following optimization:

$$\max_{C \in C_0^N} \left[\int_C [w_1 \nabla(s) + w_2 \lambda(s)] ds - w_3 \int_C (d\theta/ds)^2 ds \right], \quad (1)$$

where $0 < w_1, w_2, w_3 < 1$. Here, C_0^N denotes the set of all contours, C , of length N beginning at $s = 0$. The variable $\nabla(s)$ is the magnitude of an edge pixel along the contour and denotes the *strength* component of the criterion function; $\lambda(s) = 1$, if $\nabla(s)$ is greater than a chosen threshold and $\lambda(s) = 0$, otherwise, and it represents the *length* component; $d\theta/ds$ denotes the local curvature at the selected pixel, where $\theta(s)$ is the slope along the contour, and it is a measure of local roughness. To reduce complexity, the above optimization model is decomposed into n sub-optimization steps, each of which finds an edge segment of length N_i such that $N = \sum_{i=1}^n N_i$. These sub-optimization

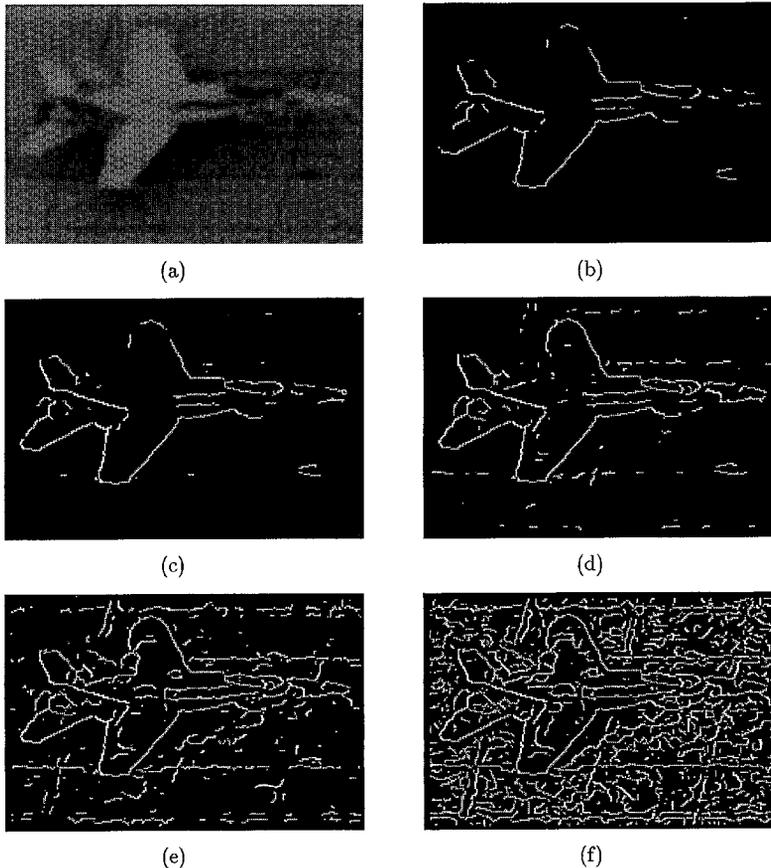


Fig. 6. Results of applying multiple thresholds (t) to an edge detector output: (a) original image; edge images obtained with (b) $t = 225$, (c) $t = 200$, (d) $t = 150$, (e) $t = 100$, (f) $t = 50$.

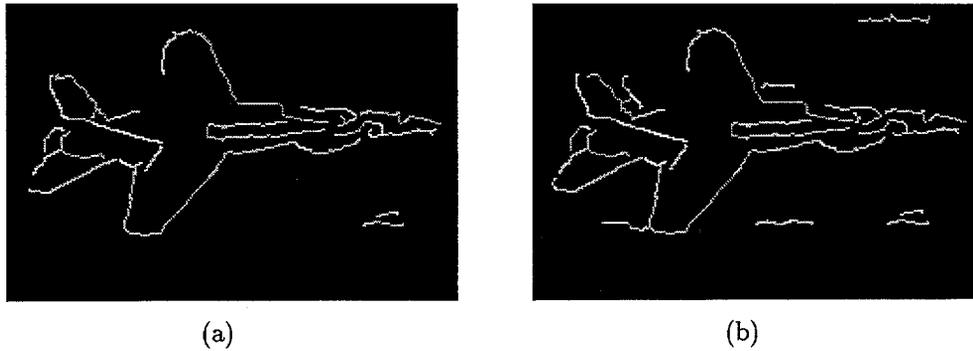


Fig. 7. Illustrating salient contour extraction: (a) detection of salient contours corresponding to edge image Fig. 6(b), (b) detection of salient contours corresponding to edge image Fig. 6(c).

steps differ in the threshold of edge magnitude (hence the use of n thresholds) which affects the $\lambda(s)$ term of equation 1.

At the first sub-optimization step, edge segment-following is initiated at a terminal pixel (one which has a single neighboring edge pixel) in the corresponding edge image [e.g. Fig. 6(b)]. To continue edge-following, a neighbor of the last selected edge segment pixel is chosen that maximizes equation (1). To account for noise, our approach allows a gap length of up to two pixels in the edge segment. Once it is terminated within the current edge image, the process is continued (i.e. the next sub-optimization step) in the image obtained with the next lower threshold [e.g. Fig. 6(c)]. The optimization process ends when the current edge image is the last of the edge-image set. Repetition of the optimization process for different starting pixels in the same initial edge image yields different contours having the same degree of saliency [e.g. contours belonging to Fig. 7(a)], while contours obtained with different initial edge images are said to have different saliency [e.g. contours belonging to Fig. 7(a) and (b)]. The contours having the same saliency are handed over to the primitive feature extraction process simultaneously, starting with the top-level configuration, i.e. contours whose initial edge image is obtained with the highest threshold [e.g. contours belonging to Fig. 7(a)]. This approach of extracting salient contours from images has been applied to and should prove to be useful for other applications, particularly when an object of interest has enough background clutter for a local feature detector to perform unsatisfactorily.

The extraction of primitive features from regulated edge contours utilizes edge-based representation of shapes. Since this representation is derived from polygonal approximations of B-spline CAD models, the primitive features comprise linear segments. In our system, we have implemented a line extraction algorithm similar to the one proposed by Lowe.⁽⁸⁾

4.2. Region feature extraction

Given the input ROI image, regions corresponding to foreground objects are identified prior to feature extraction. The region segmentation process uses gray scale intensities together with edge information (magnitude and orientation) to extract image regions. It 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.⁽¹²⁾ The joint relaxation provides edge and gray value interactions in the initial label (probability) assignment of each pixel. Only the edge orientation, and not the magnitude, is updated in each iteration. At the end of each iteration, the coincidence of edge and border values is determined. Requirement for a high degree of coincidence is necessary to obtain precise and accurate segmentation boundaries.

An important 2-D feature that is derived from an object region is the *dominant axis* characterizing the shape of the object. It is the projection of the axis of the SHGC representation of a CAD model of the object. A generic aircraft CAD model, whose sub-parts are represented by individual SHGCs, exhibits multiple dominant axes in a plane projection. In our approach, the potential dominant axes of the generic aircraft shape are obtained by connecting the extremities of a labelled region within a segmented image. To determine the extreme points, the smallest convex polygon surrounding the object region is found. It can be shown that the vertices of this polygon lie close to the local extrema of curvature points along the boundary of the labelled region. For multiple local extrema, nearby points are grouped into clusters and the cluster centers are chosen to represent the region extremities.

4.3. Nonshadow symbolic feature extraction

Assuming that any arbitrarily shaped shadow boundary can be locally represented by straight lines, our algorithm to detect potential shadow lines is

based on the test of bimodality of the local histogram. Initially, region segmentation⁽¹²⁾ is carried out within a window on either side of an extracted line and the largest region is retained. The most significant modes of the histograms of these two regions are then subjected to the bimodality test. If the separation between the modes is less than a threshold or the smaller of the two is greater than another threshold, the line is ignored. Otherwise, it is marked as a potential shadow line.

In order to separate the shadow lines from the shadow-making ones, the algorithm obtains convex groups of lines. This is a two-pass process. During the *first* pass, the entire set of lines is decomposed into subsets based on proximity and collinearity, such that lines in a subset satisfy convexity criteria when considered in pairs. This step also results in lines which fail the convexity test and are subsequently put in a pool. The *second* pass considers whether an isolated line from the pool can be put in a convex subset (subject to the convexity test) with relaxed proximity condition. The shadow and the shadow-making lines tend to form separate convex groups due to proximity of like physical features.

For each marked shadow line in a convex group, a corresponding shadow-making line from another group is sought by searching in a direction towards (or away from) the projection of the illumination point, i.e. the sun. The latter is determined from the ancillary data about the camera-platform position/orientation and the sun position together with the imaging parameters. The matching score of a pair of shadow-shadow making lines is computed from the degree of overlap of the two lines in the predicted direction. All the candidate matches of a selected shadow line are arranged according to the matching scores and marked with the corresponding group identifier. This entire matching process is repeated for all other shadow lines in that particular group. The most promising matching group is determined from the group identifiers of the candidate matches. Each line in the selected group is assigned a unique match from the candidate group based on the matching scores and enforcing similarity of spatial ordering of the selected lines and their matches. If most of the

lines in the selected group have been assigned unique matches, then the group as a whole is marked as a shadow group and the matching group is marked as a shadow-making (i.e. nonshadow) group.

Convex groups of nonshadow lines are used to derive the symbolic features of the generic aircraft class. Such features include trapezoid-like shapes for wings, tails, and rudder, and wedge-like shape for the nose part. To identify the trapezoid-like shapes, groups of three (partially closed contour) and four (fully closed contour) lines are considered. Pairwise intersections of lines are verified to occur near detected corners. Nonoverlapping line pairs which are far apart are prevented to have a high collinearity value by enforcing the condition that the average separation between the lines of a pair be proportional to the smaller line length. To overcome the problem due to oversegmentation, i.e. fragmentation of long lines into smaller parts, candidate groups of lines are merged based on collinearity measures. If the total length of the lines in a group is smaller than a certain fraction, T_{peri} of the perimeter of the trapezoid-like shape obtained by connecting these lines, then that group is discarded.

It is important to note that a range of threshold values is associated with each perceptual measure of parallelism, proximity, and collinearity. Initially, the threshold values— T_p , T_c , and T_ρ for parallelism, collinearity, and proximity, respectively—are set to the maxima of the corresponding ranges. However, these values can be relaxed during symbolic feature extraction based on the flow of evidence when multiple mutually supporting hypotheses interact (described next).

4.4. Evidence-based model matching

Once the symbolic features have been derived (see Sections 3.3 and 4.3), these need to be matched to the generic aircraft model through an evidence accumulation process which determines the number of positive evidences in support of a hypothesis. (In this paper, we do not consider negative evidences for a hypothesized object.)

During the matching process, the supervisory control module (see Fig. 5) checks each hypothesis to

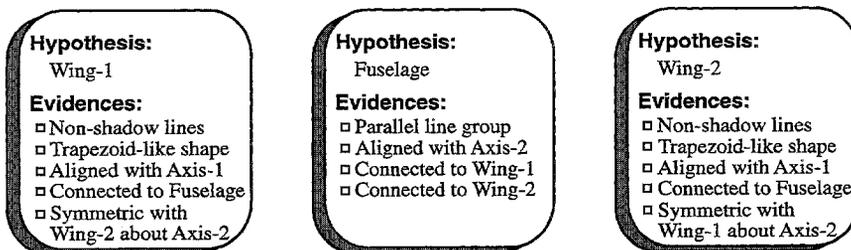


Fig. 8. Interacting hypotheses for the key structural sub-parts of a generic aircraft. The evidences (positive) are listed in the order they are sought while verifying the corresponding hypotheses.

determine its combined support based on the evidences associated with it. The combined support may be low if the evidence body is incomplete. Some pieces of evidence, known as the *critical evidences*, contribute higher support than others and must be present to accept a hypothesis. Examples of such evidences (see Fig. 8) are trapezoid-like shape, alignment with a dominant axis, and interconnection with the fuselage for the generic aircraft-wing hypothesis. During run time, the supervisory controller determines which critical evidences are missing. Typically, this situation is caused by insufficient data, either due to the screening of features at the regulator level or due to the constraints on the imaging process for that particular viewpoint. In the former situation, the controller interacts with the low-level feature regulator so that less salient features may now be available along with the existing ones. Constraints due to the viewpoint warrants the access of the qualitative database by the supervisory controller to obtain further knowledge about the alternate evidences for this hypothesis.

Further refinement of the detected aircraft shape is required to improve upon the extracted symbolic information. This usually involves completing the generic aircraft description by accounting for the missing elements of the symbolic features. This is followed by obtaining a skeleton of the refined shape which is composed of the axes of symmetry of the structural subparts. The skeleton can be directly used for mensuration purposes when performing quantitative matching.⁽¹⁰⁾ The final output consists of the identified symbolic parts of the generic aircraft.

The labelled symbolic parts are next used to direct image-based search for more localized model features. Availability of these features at progressively lower levels of the database hierarchy allows more precise classification of the recognized generic aircraft. The refocused matching process may utilize the symbolic/primitive features which have not been utilized in the generic aircraft recognition step or may request new or less-salient primitive features. Currently, our algorithm handles only qualitative model features.

5. EXPERIMENTAL RESULTS

The aircraft recognition system described in this paper is implemented on Sun Sparcstation. The information about the software system, written in C language, is detailed in Table 2. The system consists of 13 major modules. There are a total of 89 parameters, 79 fixed and 10 adjustable, that are used in these modules. A UNIX shell-level program controls the entire system. Each run of the system, as described in this section, involves processing of a single region of interest (ROI) provided as an argument to the shell program. These ROIs are obtained from photographs analyzed by the multiresolution focusing algorithm.⁽¹¹⁾ The latter is also a C code, but outside the

Table 2. Organization of the qualitative aircraft recognition system into software modules

Module	Approx. code size (#lines)	Parameters	
		Fixed	Adjustable
Region segmentation	1,500	0	5
Dominant axes	3,600	3	0
Extraction			
Saliency-based feature regulation			
Edge extraction	500	2	1
Edge linking	600	5	1
Primitive feature extraction			
Line	900	5	1
Corner	800	3	1
Convex grouping	2,300	6	0
shadow detection			
Line extraction	2,100	3	1
Line verification	2,400	3	0
Symbolic feature extraction			
Trapezoid	3,800	15	0
Aircraft features	7,100	9	0
Symbolic feature refinement	1,900	11	0
Model matching	1,800	14	0
Total	29,300	79	10

control of the shell program. The controller processes each ROI image autonomously and reports the intermediate results, including errors. In the event of premature termination of processing, such as file I/O error, the controller resumes at the point of termination after the system is restarted. Ancillary data about camera-platform position/orientation, weather condition (sunny/cloudy/hazy), sun angle, and camera parameters for the individual photographs are provided in external files. The hierarchical model database for the reported results has three levels: generic, intermediate, specific. The intermediate level consists of three categories based on engine location—large (on-wing engine), medium (on-wing/on-fuselage engine), small (concealed engine).

5.1. Data and design of experiments

The experimental results of qualitative aircraft recognition are presented using ROIs extracted from large aerial photographs. These ROIs are shown in Fig. 9. They are ordered according to increasing level of recognition complexity determined by three factors: contrast, clutter, and shadow. Such factors can assume the following qualitative values: *contrast*—low (<25%), medium-low (25–50%), medium-high (50–75%), high (>75%), where % refers to percentage of object boundary which can be successfully segmented; *clutter* (due to markings on runway and camouflage) — sparse (<25%), medium-sparse

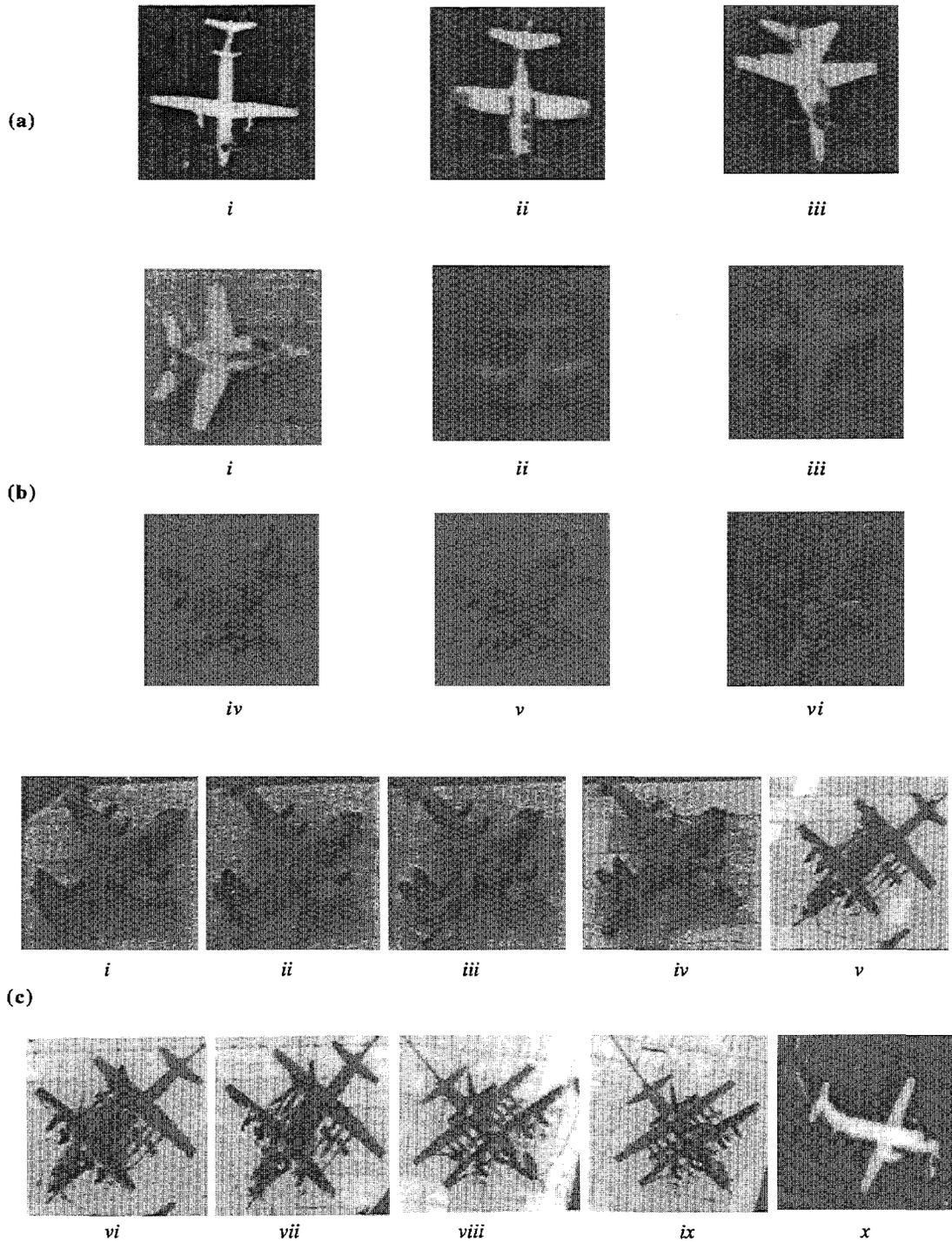


Fig. 9. Sample regions of interest extracted from large aerial images: (a) images of medium-high contrast, with sparse clutter and no shadows, (b) images of low contrast, with sparse clutter and no shadows, (c) images of medium-high contrast, with sparse clutter and shadows, (d) images of low contrast, with medium-dense clutter and no shadows, (e) images of low contrast, with medium-sparse clutter and shadows, (f) images of low contrast, with medium-sparse clutter and shadows.

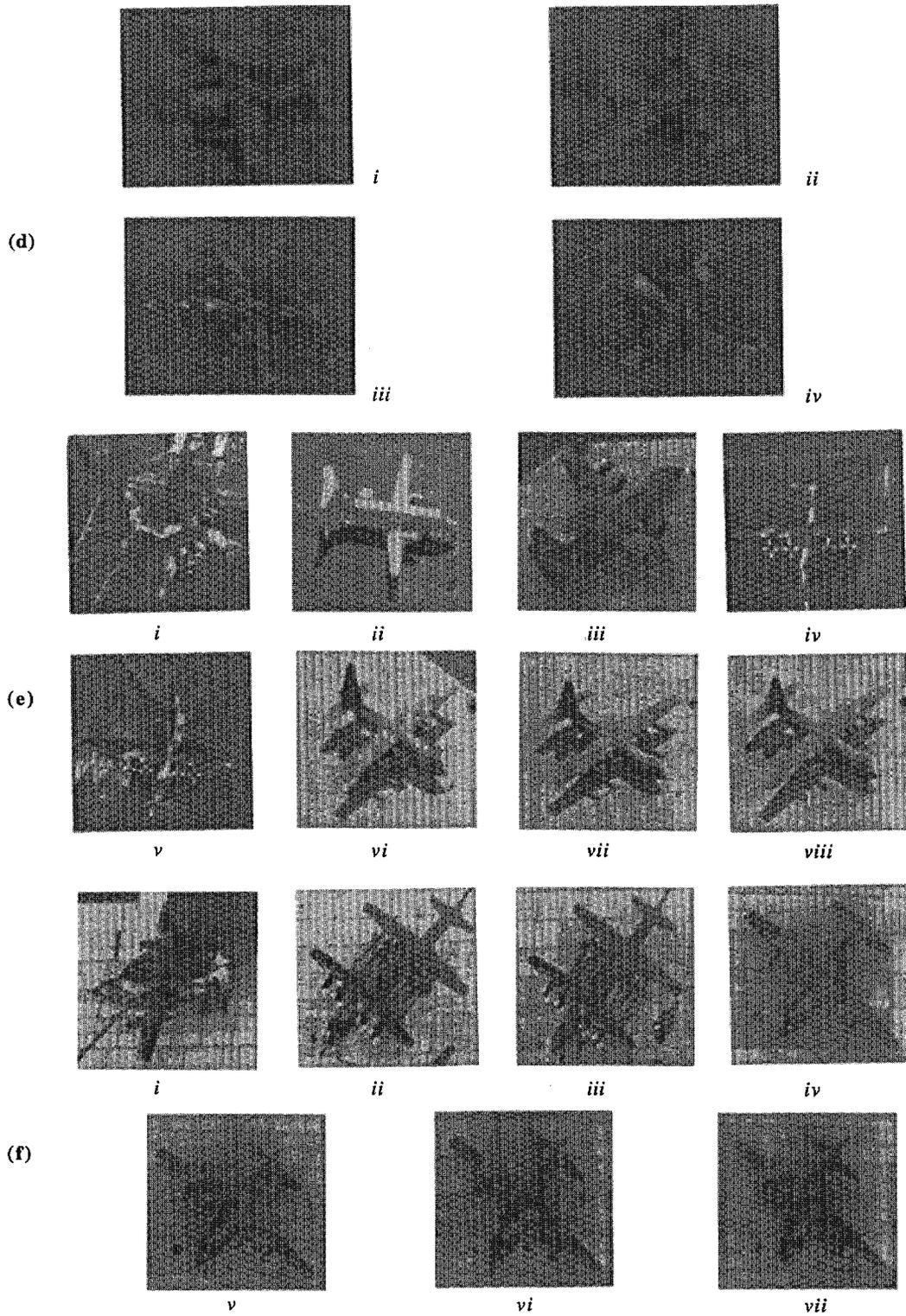


Fig. 9. (Continued)

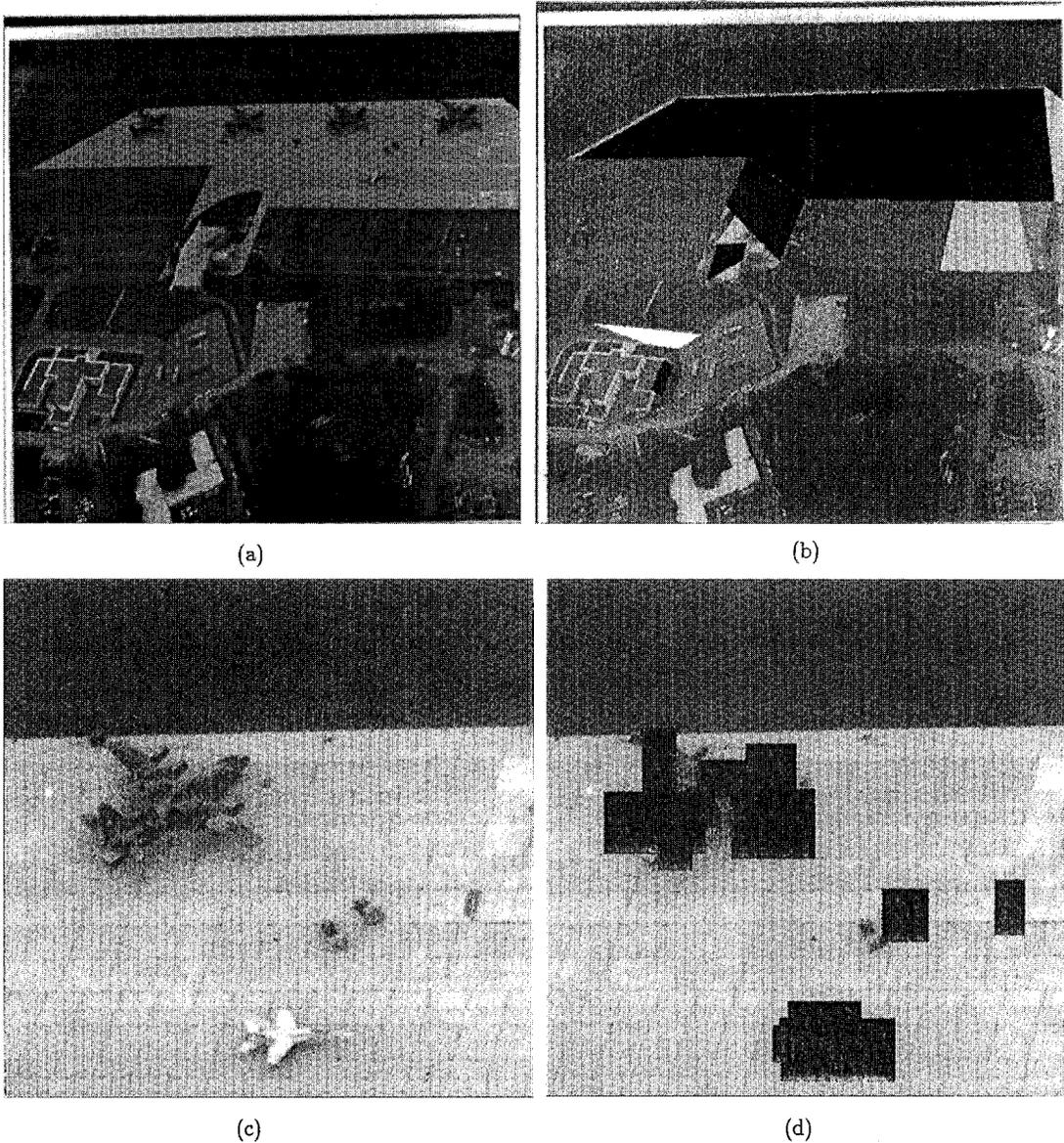


Fig. 10. An aerial view of an airfield: (a) Original image ($4K \times 4K$), (b) Preliminary regions of interest (ROIs, black regions) in (a), (c) A close-up of the preliminary ROIs of (b), (d) New ROIs found in (c).

(25–50%), medium-dense (50–75%), dense ($>75\%$), where % refers to percentage of nontarget features within a given neighborhood of a target feature; *shadow* — diffused ($<25\%$), medium (25–75%), strong ($>75\%$), where % refers to percentage of shadow boundary which can be successfully segmented. Rather than developing the statistics of performance on very similar images, we have selected three sample images from the ROIs of Fig. 9 that a system like ours will encounter in a practical application. We show detailed results of evaluation of the various components of our aircraft recognition system using these selected images. Our analyses are expressed in terms of hit-ratio, i.e. ratio of number of target features found and total number of features

(target/clutter) found, and miss-ratio, i.e. ratio of number of target features not found and actual number of target features present. Note that, by definition, hit- and miss-ratios are not complement of each other, i.e. $\text{hit-ratio} \neq 1 - \text{miss-ratio}$.

5.2. Examples of varying complexities

Example 1. Figure 10(a) shows an aerial photograph ($4K \times 4K$) in which several ROIs, displayed in Fig. 10(b)–(d), are identified using the multiresolution focusing approach. 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. 10(d). The output of the salient

contour detection step for this ROI (see Fig. 6) has already been presented in Fig. 7. In our implementation, we have selected the five threshold (t) values indicated in Fig. 6 which are fixed for *all* images. Figure 7(a) shows the top-level configuration, i.e. most salient contours, consisting of the aircraft in this case.

The following step is to extract the primitive features from this contour set. The result of line fitting to the salient structures is shown in Fig. 11(a). Segmented regions and the dominant axes of regions are shown in Fig. 11(b). Figure 12(a) shows the six convex sets of lines identified using convex grouping procedure.

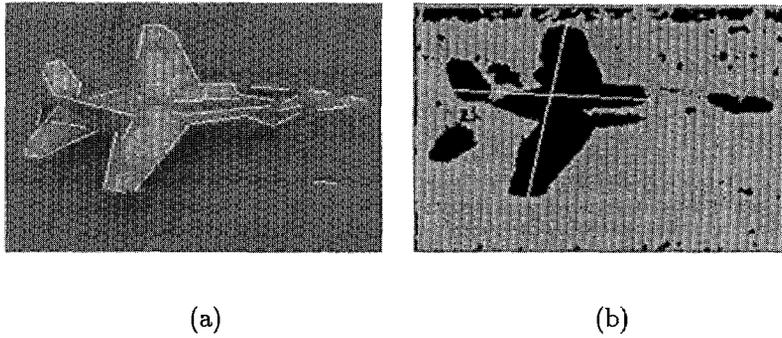
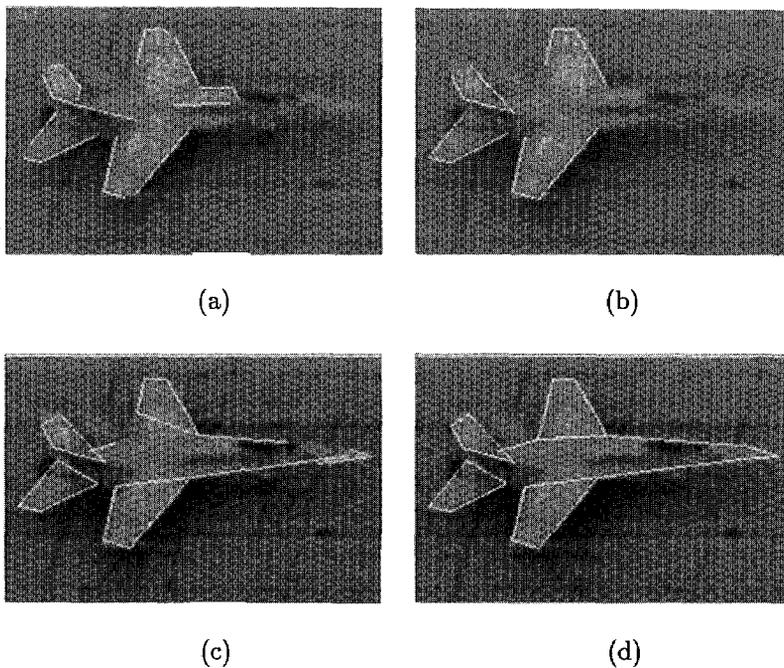


Fig. 11. Results of low-level processing of the bottom ROI in Fig. 10(d): (a) fitting straight lines to the contours of Fig. 7(a), (b) segmented regions and extracted dominant axes for the largest foreground region.



Total Engines found on the Left Wing: 0
 Total Engines found on the Right Wing: 0
 Total Engines found on the Left Rear Fuselage: 0
 Aircraft Class: << small aircraft >>

(e)

Fig. 12. Results of qualitative object recognition: (a) six convex groups of lines identified in Fig. 11(a), (b) trapezoid-like shapes identified using these groups, (c) structural parts found during generic object recognition, (d) refined structural parts that are also labelled, (e) class recognition.

These are used to extract trapezoid-like features shown in Fig. 12(b) which are the symbolic descriptions of some of the subparts, like wings, tails, and rudder (see Fig. 4).

During the generic object recognition step, the order of evidence accumulation is for the wings first, followed by that for the fuselage, with the nose being the last. 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. 4 have been satisfied, can their ensemble be recognized as a generic aircraft. The identified subparts are shown in Fig. 12(c). Inability to identify both wings or the fuselage is considered to be a recognition failure for the generic aircraft class. In this case, both wings are detected and so also the fuselage, hence, the recognition of a generic aircraft is successful. The connectivity information of the parts is exploited to obtain more complete descriptions of the

subparts as shown in Fig. 12(d). Next, an improved classification of the generic aircraft is sought based on the engine location (Fig. 4). 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. 12(e)].

The significance of salient feature selection is summarized in Table 4. Although, a traditional feature extraction process will result in undersegmentation (e.g. 18 lines for $t = 225$) or oversegmentation (e.g. 213 lines for $t = 50$), the number of lines from the most salient structure [shown in Fig. 7(a)] is 36. The set of lines belonging to the most salient structures constitutes only 17% of the total lines obtained from the edge image of Fig. 6(f). At the same time, the salient contour set accounts for nearly 84% of the “useful” lines associated with the aircraft. This percentage is referred to as the “ground truth” in Table 4. It expresses a significant gain in terms of computational

Table 3. The factors which affect the feature extraction processes and, hence, determine the level of difficulty in categorizing the target object, i.e. a generic aircraft and its class, in the chosen examples

Factors	Example 1	Example 2	Example 3
Contrast	Medium-low (target/background)	Medium-high (target/background), high (target/shadow) high (shadow/background)	Medium-low (target/background), low (target/shadow) low (shadow/background)
Clutter Shadow	Medium-sparse Diffused	Sparse Strong	Medium-dense Strong

Table 4. Significance of saliency-based feature regulation for the ROI of Example 1

No. of lines detected using edge magnitude threshold of					No. of lines from from most salient structures (l_1)	Ratio l_1/L_5	Fraction of aircraft contour lines in l_1 (ground truth)
$t = 225$ (L_1)	$t = 200$ (L_2)	$t = 150$ (L_3)	$t = 100$ (L_4)	$t = 50$ (L_5)			
18	28	40	91	213	36	0.17	0.84

The lines referred to in this table are those that are at least 10 pixels long.

Table 5. Significance of symbolic feature extraction for the ROI of Example 1

No. of lines from most salient structures (l_1)	No. of distinct lines forming convex groups (l_2)	No. of distinct lines forming trapezoids (l_3)	No. of additional lines needed for recognition (l_4)	Ratio l_2/l_1	Ratio $(l_3 + l_4) / l_1$
36	18	13	5	0.5	0.5

The lines referred to in this table are those that are at least 10 pixels long.

efficiency for any subsequent model-matching step. The computational advantage gained because of the use of symbolic features is evident from the results of Table 5. The overall effect of salient structure determination (17% of the original set of lines retained)

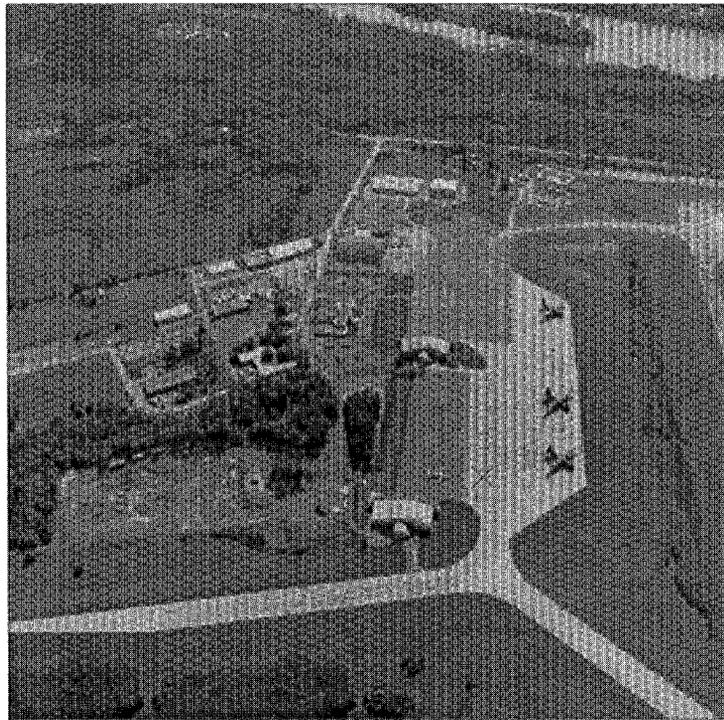
and symbolic feature grouping (50% of the salient lines used) is the retention of only 8.5% of the lines obtained from Fig. 6(f) for aircraft recognition. Table 6 presents the performance results of the feature extraction processes — salient edge, region, non-shadow edge, and symbolic — in terms of hit- and miss-ratios defined earlier. The merit figures follow the experimental results discussed above, except for nonshadow feature extraction which is not applicable (N/A) to this example. A hit ratio of 1.0 for the symbolic feature extraction process implies that no nonaircraft feature (trapezoid-shaped) has been extracted.

Table 6. Performance evaluation of various feature extraction processes in Example 1 under the conditions of Table 3

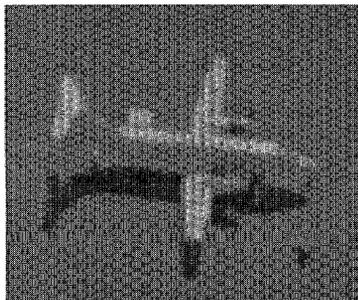
Salient edge features		Region features		Non-shadow edge features		Symbolic features	
<i>H</i>	<i>M</i>	<i>H</i>	<i>M</i>	<i>H</i>	<i>M</i>	<i>H</i>	<i>M</i>
0.84	0.10	0.75	0.20	N/A	N/A	1.0	0.20

H and *M* represent hit- and miss-ratios, respectively.

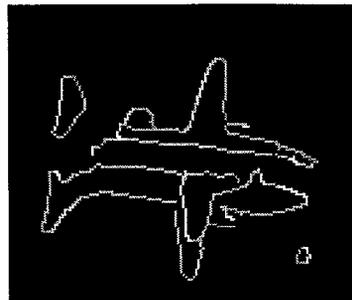
Example 2. A second aerial image is shown in Fig. 13(a). In this example, we present the results of



(a)



(b)



(c)

Fig. 13. 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.

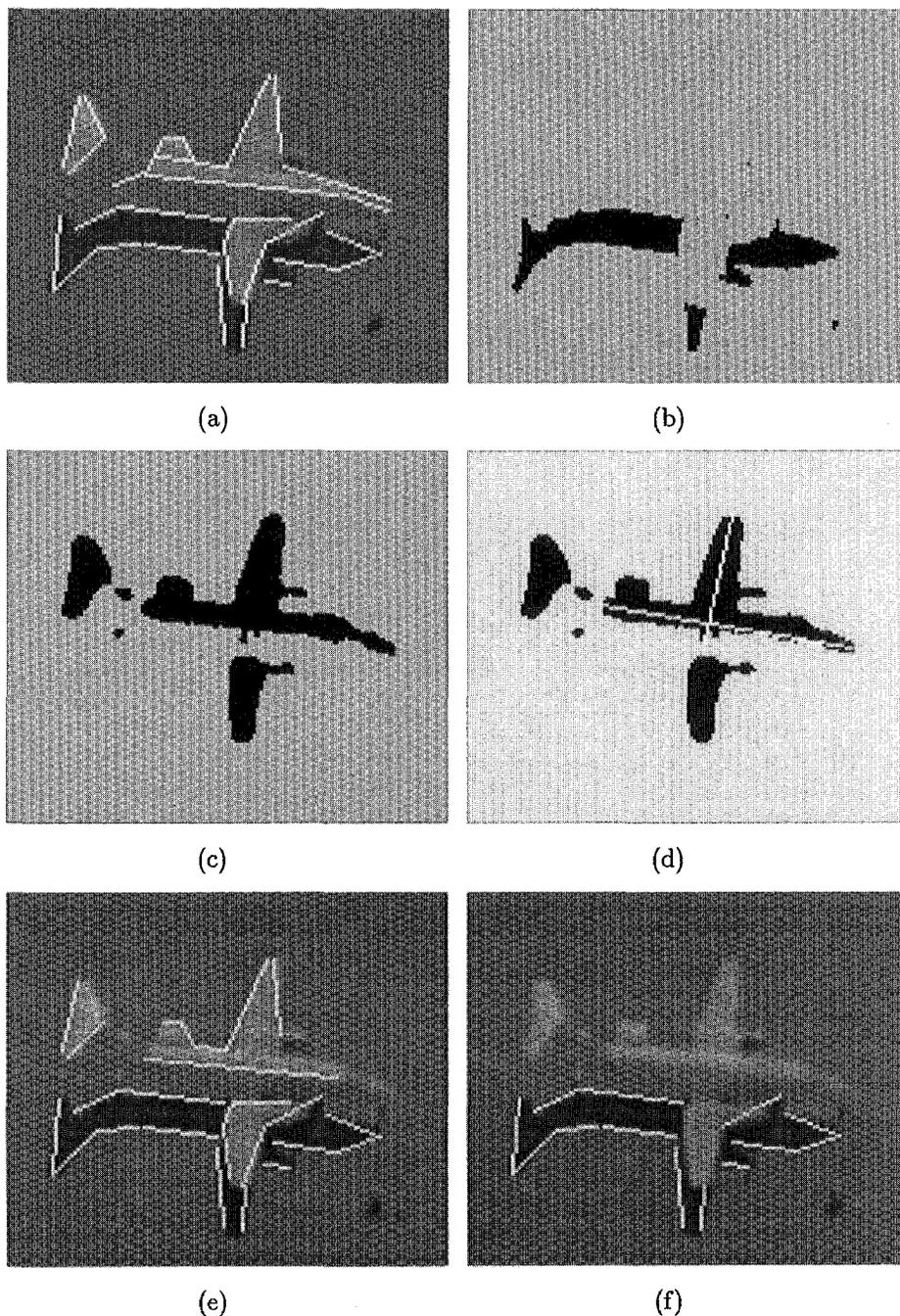


Fig. 14. Results of feature extraction: (a) straight lines fitted to the most salient contours, (b) segmented regions where 0 = shadow and 1 = background, (c) regions where 0 = object and 1 = background, (d) dominant axes for the largest object region where axis-1 = wing axis, axis-2 = fuselage axis, (e) potential shadow lines, (f) resolved shadow lines.

analyzing the ROI of Fig. 13(b). The top-level (most salient) configuration of perceptually salient contours is shown in Fig. 13(c). The linear segments of this structure are displayed in Fig. 14(a). The ROI is segmented into three sets — shadow [Fig. 14(b)], object, and background [Fig. 14(c)]—using recursive ap-

plication of the segmentation algorithm. The dominant axes of the largest object region are shown in Fig. 14(d). The potential shadow lines based on the bimodality test of neighborhood histograms are displayed in Fig. 14(e). To identify the nonshadow lines, the illumination point projection is first determined

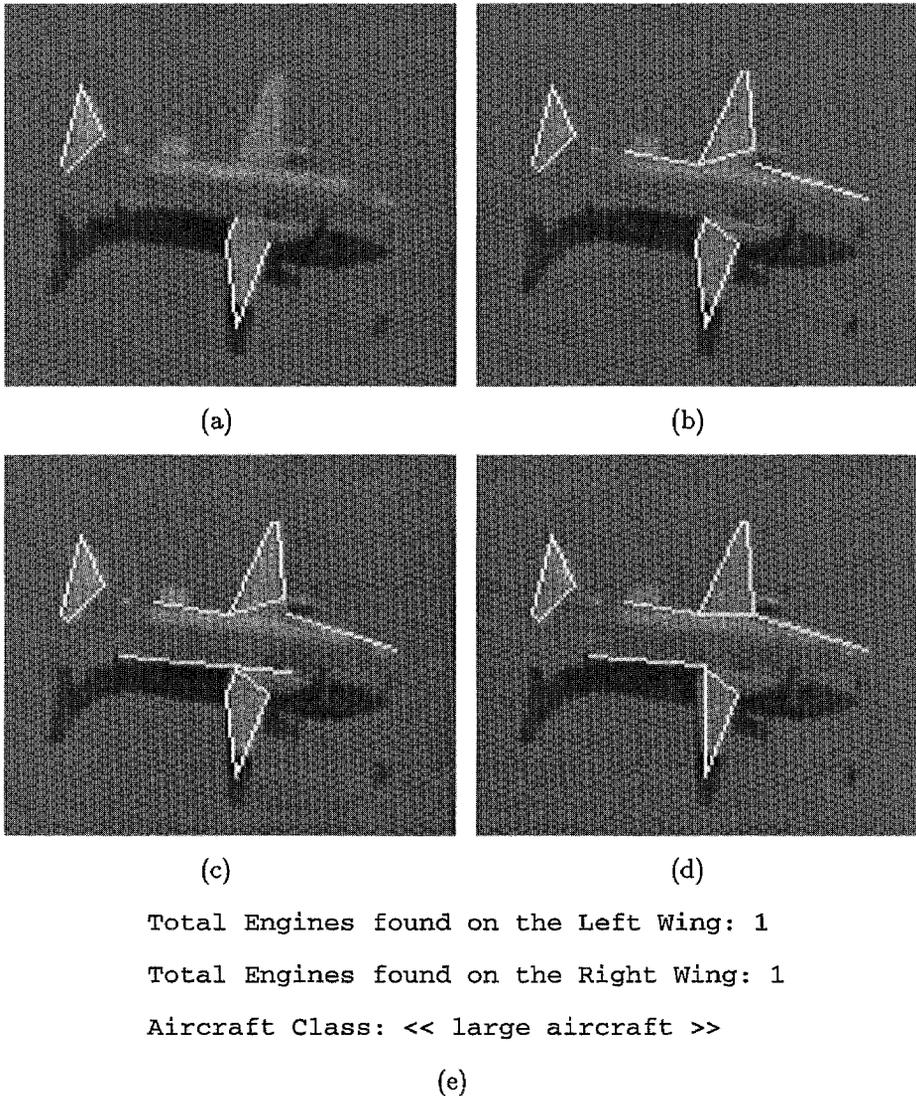


Fig. 15. Results of qualitative object recognition: (a) trapezoid-like shapes identified using nonshadow groups, (b) symbolic features recognized with low confidence, (c) improved recognition using regulated features, (d) refined structural parts, (e) class recognition.

using the ancillary data. The coordinates of the camera-platform position are (527, 337, 560) m and that of the point of the intersection of the line-of-sight (LOS) with the ground are (−500, −290, −560) m; the roll of the camera about the LOS is 20°. The sun position is noted to be *in front* of the camera. The shadow lines which paired up with “nonshadow” lines are displayed in Fig. 14(f). Next, convex groups of the nonshadow lines are used to extract trapezoid-like symbolic features of Fig. 15(a).

Figure 15(b) shows the results of initial recognition during which only one trapezoid-like feature is available to support a wing concept (for the right wing) due to its alignment with the wing axis (axis-1). Additional evidence for this hypothesized wing is obtained from

a successful search for elongated blobs representing engine. Next, a search region for the second wing is set up along axis-1 on the other side (away from the hypothesized wing) of the fuselage axis (axis-2). A line pair is located within this region that aligns with axis-1 and for which the presence of engine can also be verified. Thus, the symbolic feature and the line pairs are retained as candidates for the two wings as shown in Fig. 15(b).

The supervisory control now interacts with the feature regulator to identify candidate lines for the missing parts of the fuselage in the image region specified by the two hypothesized wings and the hypothesized fuselage subpart. The “shadow” lines removed earlier are considered first, since these are

perceptually more salient than any other “non-shadow” line that may be obtained from the next set of less salient contours. Two such lines are found which made up for the missing bottom part of the fuselage. The final results are shown in Fig. 15(c) when an aircraft is said to be have been recognized. The symbolic parts are further refined as shown in

Table 7. Performance evaluation of various feature extraction processes in Example 2 under the conditions of Table 3

Salient edge features		Region features		Non-shadow edge features		Symbolic features	
<i>H</i>	<i>M</i>	<i>H</i>	<i>M</i>	<i>H</i>	<i>M</i>	<i>H</i>	<i>M</i>
0.54	0.14	0.98	0.35	1.0	0.14	1.0	0.20

H and *M* represent hit- and miss-ratios, respectively.

Fig. 15(d). Finally, the class of the generic aircraft is determined to be *large* based on engine locations as seen in Fig. 15(e).

The performance results of the feature extraction processes are summarized in Table 7. The relatively low value of the hit-ratio in the case of salient edge feature extraction process is attributable to the contours of the shadow cast by the aircraft. The self-shadowing of the aircraft seen in Fig. 13(b) is also the cause of the higher miss-ratio in the case of region feature extraction.

Example 3. The third aerial image is shown in Fig. 16(a) of which Fig. 16(b) constitutes an ROI. The segmentation result and the extracted dominant axes are shown in Fig. 16(c). The illumination point projection is determined using the ancillary data. The coordinates of the camera-platform position are

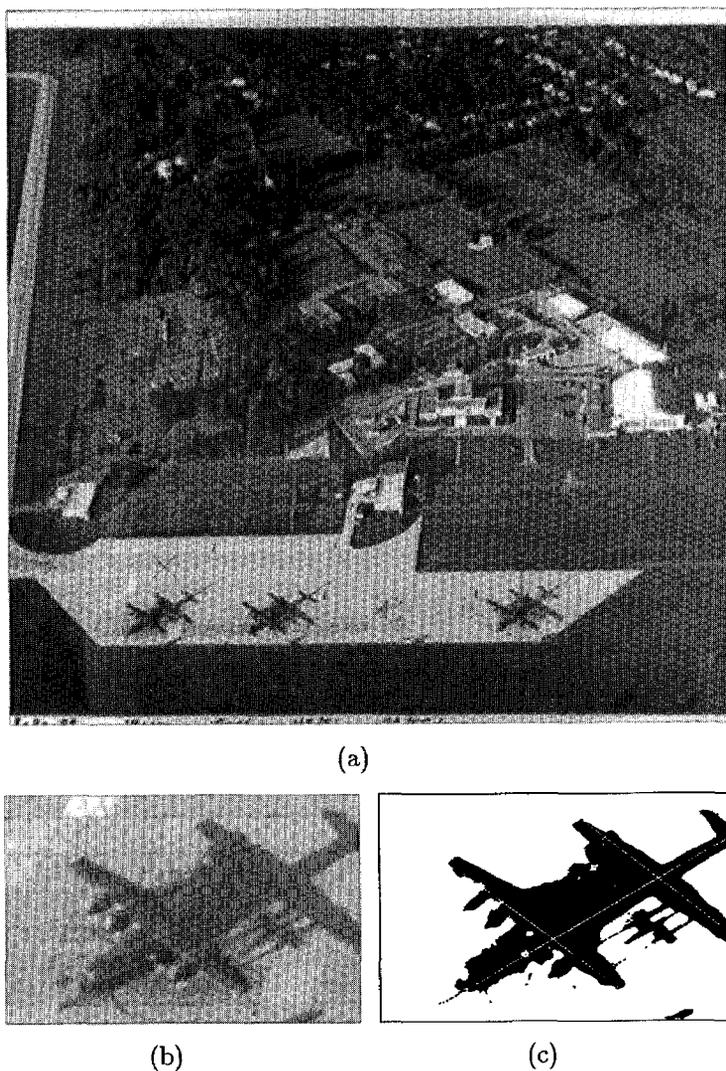


Fig. 16. A second aerial image: (a) original image ($4K \times 4K$), (b) a ROI image (300×450) of the aircraft marked with a \times in (a), (c) extracted dominant axes.

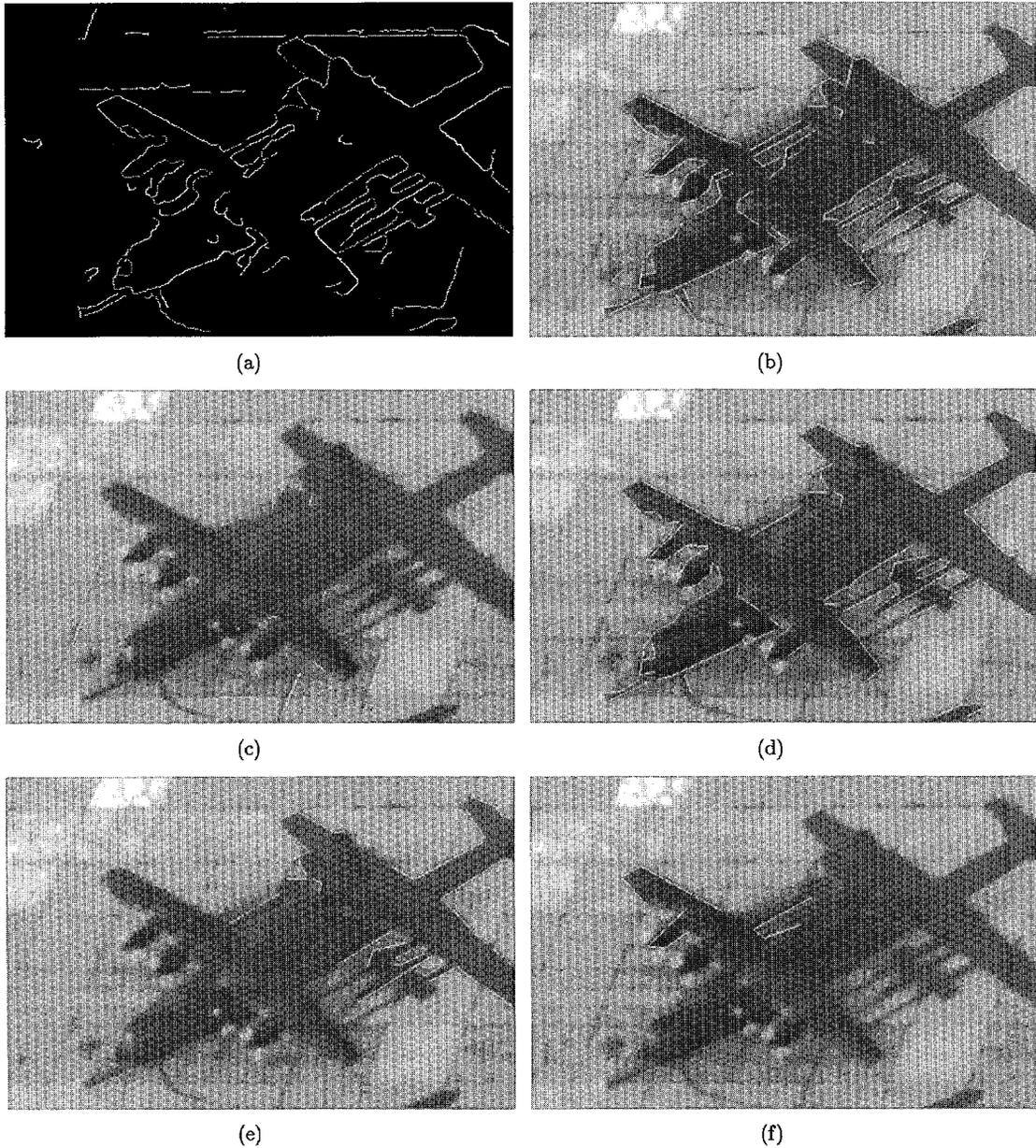


Fig. 17. Results of feature extraction: (a) detection of most salient structure, (b) straight lines fit to the structure of (a) superimposed on the ROI image, (c) lines fit to the next incremental salient structure, (d) potential shadow lines, (e) resolved shadow lines, (f) trapezoid-like shapes identified using nonshadow groups.

(-984, -115, 700) m, that of the point of the intersection of the LOS with the ground are (1150, 150, -700) m, and the roll of the camera about the LOS is -21° . The sun position is recorded as *behind* the camera.

Figure 17(a) shows the top-level structure. The line-fits to this structure and the next incremental salient structure are displayed in Fig. 17(b) and (c). Once again, the computational advantage gained by using the salient structure detection step is indicated in

Table 8. Here, the salient structures constitute only 26% of the lines obtained from the edge image corresponding to the lowest edge threshold. Table 8 also illustrates the difficulty posed by shadows as nearly equal number of lines (the ground truth) belong to the actual aircraft contour (38% of the salient structure lines) and the shadows (35% of the salient structure lines). Convex groups are formed using the lines of Fig. 17(b). The potential shadow lines identified among the lines of Fig. 17(b) are shown in Fig. 17(d).

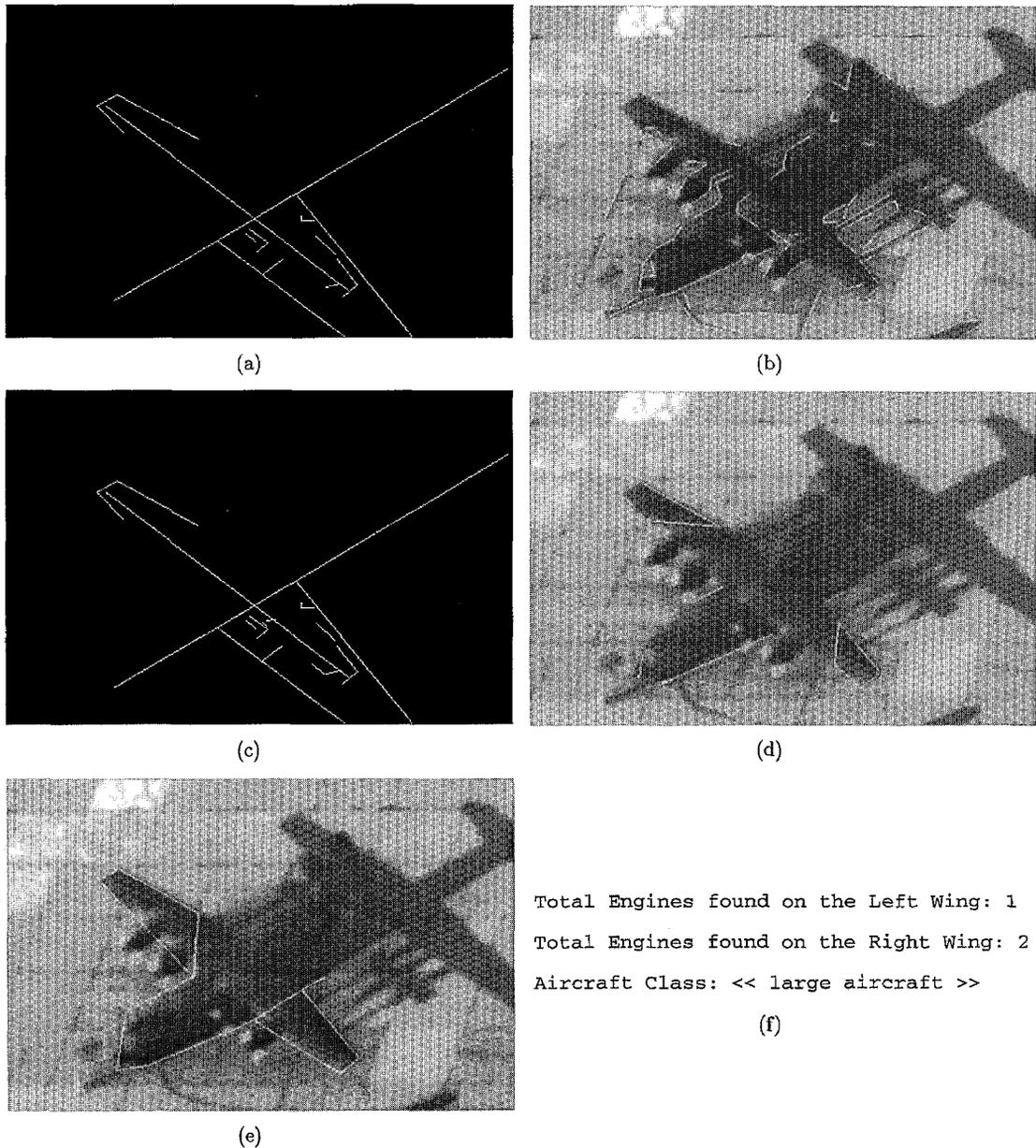


Fig. 18. Results of qualitative object recognition: (a) a hypothesized wing and a search region for the second wing, (b) current nonshadow lines together with those of Fig. 17(c), (c) additional nonshadow lines within the search region of (a), (d) recognition after verification of the second wing, (e) refined structural parts, (f) class recognition.

To resolve the nonshadow lines, the rightmost of the two nearly parallel axes of Fig. 16(c) is determined to be due to shadow based on the illumination projection point location. The final shadow lines are displayed in Fig. 17(e). The groups of nonshadow lines are next used to obtain the trapezoid-like symbolic features of Fig. 17(f).

One of the symbolic features of Fig. 17(f) that aligned with one of the dominant axes, axis-1, is

hypothesized as the wing, wing-1 (say). This is shown in Fig. 18(a). Further evidence of wing is obtained from a successful search for engine feature in the vicinity of wing-1. To obtain evidence of the other wing, wing-2 (say), a search region is set up as shown in Fig. 18(a) based on the image location of wing-1 and the condition of symmetry of the wings about the fuselage axis. However, the nonshadow lines contained within this region fail to identify any symbolic

Table 8. Significance of shadow detection and symbolic feature extraction for the ROI of Example 3

No. of lines from the most salient structures (l_1)	Fraction of lines from E_{50} in set l_1	Fraction of l_1 belonging to aircraft (ground truth)	Fraction of l_1 belonging to shadows (ground truth)	No. of shadow lines found (l_2)
150	0.26	0.38	0.35	36

No. of additional lines from the next salient structures (l_3)	No. of distinct lines forming trapezoids (l_4)	No. of additional lines needed for generic recognition (l_5)	Ratio $(l_4 + l_5)/(l_1 + l_3)$
24	6	5	0.06

The lines referred to in this table are those that are at least 10 pixels long. The edge image obtained using a magnitude threshold of $t = 50$ is denoted by E_{50} .

feature that may support a wing hypothesis. Additional nonshadow lines derived from less salient contours of Fig. 18(b) are combined with the existing nonshadow lines within the search region [Fig. 18(c)]. The initial (maximum) values of the various perceptual constraints fail to produce any meaningful perceptual grouping as the lines are few in number and are separated apart. However, since the hypotheses of wing-1 and wing-2 are mutually reinforcing, the constraints are relaxed in steps. Particularly, the lowering of the threshold of the proximity, T_p , of a line-pair intersection to a detected corner and T_{peri} (initial values of $T_p = 30.0$ and $T_{peri} = 0.9$ in steps of $\Delta T_p = 7.0$ and $\Delta T_{peri} = 0.1$) cause grouping of lines to occur. As a result, a trapezoid-like symbolic feature emerges (the left wing tip) that drives the subsequent steps of recognition shown in Fig. 18(d)–(e). Finally, the refocused matching process determines the class of this aircraft as *large* as indicated in Fig. 18(f).

Upon evaluating the processes of shadow identification and symbolic feature extraction, it is seen (Table 8) that the number of useful lines for aircraft recognition is reduced to only 6% of the total number of lines which would have to be considered otherwise. Performance results of the four key feature extraction processes are included in Table 9. The numbers indicate the difficulty posed by the low contrast and the clutter of the image for these processes, i.e. smaller values of the hit-ratio and larger values of the miss-ratio.

Evaluation of results. Figures 18 and 19 illustrate the overall performance of the feature extraction processes for the examples considered in this section. The performance is measured under two of the conditions identified in Table 2: contrast and clutter. The effect of

Table 9. Performance evaluation of various feature extraction processes in Example 3 under the conditions of Table 3

Salient edge features		Region features		Non-shadow edge features		Symbolic features	
H	M	H	M	H	M	H	M
0.38	0.16	0.55	0.02	0.5	0.45	0.66	0.5

H and M represent hit- and miss-ratios, respectively.

shadows on the recognition process is implicit in the contrast factor, since presence of shadows may reduce the target contrast as in Example 3. It is evident from these performance figures that our chosen examples span the possible scenarios (with the exception of dense clutter) which may be encountered in practice, viz., the variations in contrast, clutter, and shadow. The figures of merit (hit-ratio and miss-ratio) are, therefore, reasonable indication of the capabilities of the system under real-world conditions. Our system fails on the very low-contrast images of Fig. 9, such as Fig. 9(b):(vi), Fig. 9(d), and Fig. 9(e):(i), (iv), (v), due to segmentation problem which we discuss later.

5.3. Discussion of problematic issues

The current limitation of the aircraft recognition system is primarily due to image segmentation. The secondary factors include limited symbolic feature extraction, nonadaptive recognition strategy, and imprecise auxiliary data.

Image segmentation is an extremely important and difficult low-level task. The difficulty arises when the segmentation performance needs to be adapted to the changes in image quality that is affected by variations in environmental conditions, time of day, etc., as illustrated by the example images of this paper. Despite the large number of segmentation techniques presently available, none performs adequately across all possible scenarios likely to be encountered in practice. Any technique, including the segmentation algorithm employed in this research, yields poor performance if it cannot adapt to the variations observed in real-world images as shown in this paper. Genetic and reinforcement learning-based techniques can help here.^(13,14)

The process of symbolic feature extraction in our system is based on grouping of primitive features, primarily lines and corners. Like most current grouping techniques, our system uses only perceptually motivated, low-order geometric relationships (such as collinearity, cotermination, parallelism, proximity, etc.) to assemble simple features of the same type. The usefulness of the derived symbolic features is dependent on the reliability of the primitive feature extraction step, including image segmentation. Augmenting the domain-independent perceptual grouping process with model-based grouping of high-order structural features

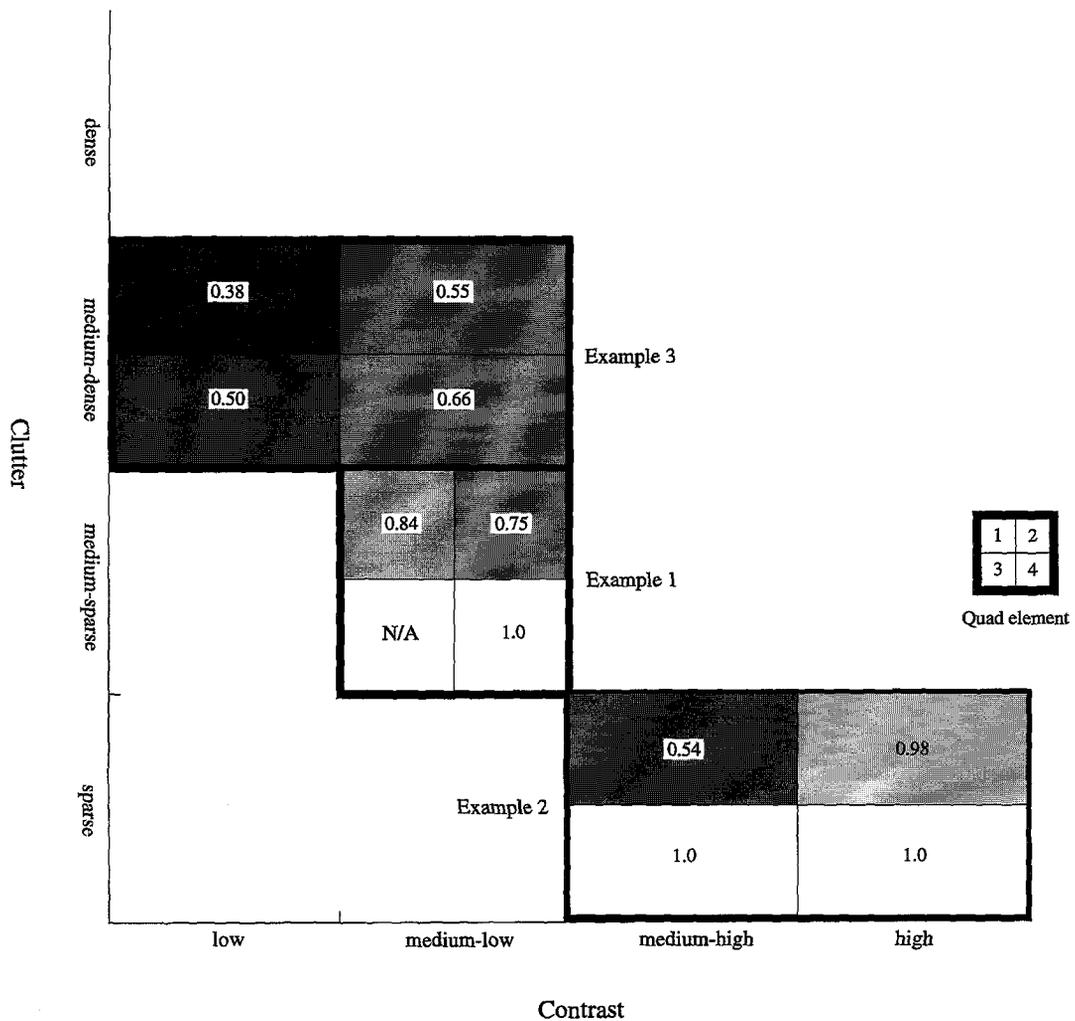


Fig. 19. Overall performance, in terms of hit-ratios, of the feature extraction processes as functions of the contrast and clutter factors. Each quad element represents the processes in the following way: 1—salient edge, 2—region, 3—nonshadow edge, 4—symbolic. The hit-ratio values range from 0 (black) to 1 (white); N/A implies that particular value is not available.

(e.g. curves and blobs) will lead to increased robustness and indexing power of the symbolic features.

The success of the final model matching step is determined by the scope of the implemented recognition strategy. The effectiveness of a rigid recognition strategy in tackling the complex problem of model matching is bound to be limited in multisenario environments. The limitation may be partially attributed to the initially unknown pose of the input instance of the model and the varying conditions (besides the camouflage, concealment, and deception) in which an object may appear in an image. Our system lacks the capability of automatic acquisition of recognition strategies, which remains a bottleneck in the development of automated IU systems applied to real-world problems.

6. CONCLUSIONS

In this paper, we have described a system for recognition of aircraft in complex, perspective aerial imagery and presented an algorithm to recognize and classify aircraft using qualitative features. Our approach is motivated by the difficulties posed by real-world scenarios, such as occlusion, shadow, cloud cover, haze, seasonal variations, clutter and various forms of image degradation. The main contributions of this research are the extraction of perceptually salient primitive features and their use in a regulated fashion, use of heterogeneous geometric and physical models associated with image formation for feature extraction and subsequent recognition, and integration of high-level recognition processes with low-level feature

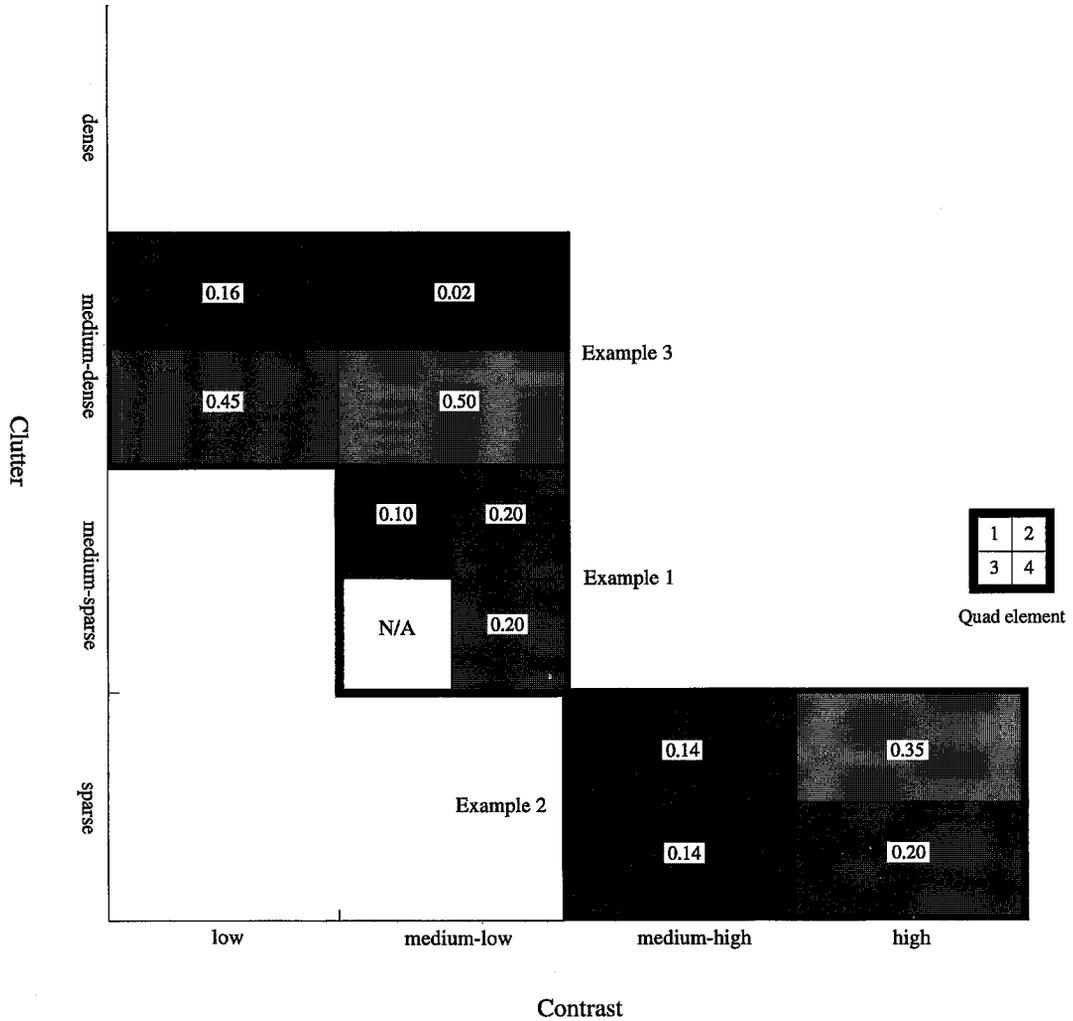


Fig. 20. Overall performance, in terms of miss-ratios, of the feature extraction processes as functions of the *contrast* and *clutter* factors. Each quad element represents the processes in the following way: 1—salient edge, 2—region, 3—nonshadow edge, 4—symbolic. The miss-ratio values range from 0 (black) to 1 (white); N/A implies that particular value is not available.

extraction ones. Real-world data, highlighting the difficulties of aircraft recognition in practical situations, are used to demonstrate the effectiveness of our proposed approach. We have provided both quantitative and qualitative figures of merit illustrating the performance of various aspects of our system. Elsewhere, we have utilized the qualitative results to search for more specific aircraft models. In its current state, the system can be operated by anyone without knowing the details of the algorithm. The intermediate results are logged for off-line analysis of the performance of the system subcomponents. Our system needs further improvement in the integration and the simultaneous hypothesis verification steps to handle scenarios like the one illustrated in Fig. 1(c). In addition, more accurate auxiliary data,

particularly those of camera/platform geometry, should facilitate robust feature extraction and hypotheses verification.

The issues and concerns related with extending the scope of our system to handle a wide variety of images in a practical setting can be addressed within the context of machine learning-based image understanding. Learning provides adaptation of algorithms to changes in environment and uses the experience to improve recognition performance.⁽¹⁵⁾ As a future step, we will integrate machine learning components into the current system to increase its scope and to improve its robustness. We invite pattern recognition and computer vision researchers to use the imagery, shown in this paper, for the development of robust aircraft recognition approaches.

7. SUMMARY

An important goal of photointerpretation (PI) is image exploitation, particularly aerial imagery. The problem of PI is one of identifying instances of "known" object models, usually geometric models, in images acquired from a platform, such as by a satellite or a reconnaissance aircraft. In real-world PI scenarios, there are factors which 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. All of these problems put heavy requirements on any image understanding (IU) system to be robust. Among the typical objects requiring identification in PI tasks, buildings and aircraft have received the maximum attention in the literature. Recognition of buildings has proved to be relatively easier than that of aircraft because of the structural simplicity of the former. Although there have been several aircraft recognition systems proposed in the past, very few of these have actually addressed the concerns of real world or are demonstrated to be effective in practical scenarios.

In this paper, we describe an IU system for aircraft recognition that addresses some of the issues related to geometric model-based object recognition and also the variabilities of real-world scenarios, such as shadow, clutter, and low contrast. Our system uses a hierarchical representation, consisting of *qualitative-to-quantitative* descriptions, of object models (aircraft in this case). Such descriptions vary from symbolic features (e.g. aircraft wing) to primitive geometric entities (e.g. points, lines) and allow increasingly focused search of the precise models in the database to match the image features. The organization of our model database is in the form of a hierarchy of generic-to-specific information about generic objects (e.g. aircraft), object classes (e.g. jumbo aircraft), specific objects (e.g. Boeing 747), and aspects of an object. It is the *specificity* of the information available at any given level of the hierarchy that controls the focus of the search. Additionally, our approach emphasizes the importance of using *symbolic features* which are known to be aspect invariant in majority of the real-world perspective imagery except for extreme viewpoint situations. To account for image variabilities, our system exploits *heterogeneous models* such as those of camera/platform, sun, shadow to derive symbolic features for the objects in a robust manner. Finally, we have brought in a novel aspect to the recognition problem by *regulating* the extracted *primitive features* based on their saliency. We demonstrate that this step helps to distinguish the relevant features from the image clutter, thereby reducing the complexity of the search problem.

The *main contributions* of this research are the extraction of perceptually salient primitive features and their use in a regulated fashion, the use of heterogeneous geometric and physical models associated

with image formation for feature extraction and subsequent recognition, and the integration of high-level recognition processes with low-level feature extraction ones. Real-world data, highlighting the difficulties of aircraft recognition in practical situations, are used to demonstrate the effectiveness of our proposed approach.

Acknowledgements—This work was supported in part by a grant from Honeywell Inc., and DARPA/AFOSR grants F9620-93-1-0624 and F49620-95-1-0424.

REFERENCES

1. R. A. Brooks, Symbolic reasoning among 3-dimensional models and 2-dimensional images, *Artificial Intell.* **17**, 285–349 (1981).
2. A. Huertas and R. Nevatia, Detecting buildings in aerial images, *Complete. Vis. Graph Image Process.* **41**(2), 131–152 (1988).
3. R. B. Irvin and D. M. McKeown, Methods for exploiting the relationship between buildings and their shadows in aerial imagery, *IEEE Trans. System Man and Cybernet.* **19**(6), 1564–1575 (1989).
4. C. Bjorklund, M. Noga, E. Barrett and D. Kuan, Lockheed imaging technology research for missiles and space, *Proc. DARPA Image Understanding Workshop*, pp. 332–352. Palo Alto, California, (May 1989).
5. V. Clément, G. Giraudon, S. Houzelle and F. Sandakly, Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture, *IEEE Trans. Geosci. Remote Sensing* **31**(4), 779–791 (1993).
6. A. Sha'ashua and S. Ullman, Structural saliency: the detection of globally salient structures using a locally connected network, *Proc. IEEE 2nd Int. Conf. Comp. Vision*, pp. 321–327. Tarpon Springs, Florida (December 1988).
7. R. C. Bolles and P. Horaud, 3DPO: a three-dimensional part orientation system, *Int. J. Robot. Res.* **5**(3), 3–26 (1986).
8. D. G. Lowe, *Perceptual Organization and Visual Recognition*. Kluwer, Boston, Massachusetts (1985).
9. D. McKeown, W. A. Harvey and J. McDermott, Rule-based interpretation of aerial imagery, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**(5), 570–585 (1985).
10. S. Das, B. Bhanu and C.-C. Ho, Generic object recognition using multiple representations, *Image Vis Comput.* **14**(5), 323–338 (1996).
11. H. Nasr, B. Bhanu and S. Lee, Refocused recognition of aerial photographs at multiple resolution, *Proc. SPIE*, Vol. 1098, pp. 198–206. Orlando, Florida (March 1989).
12. B. Bhanu and R. D. Holben, Model-based segmentation of FLIR images, *IEEE Trans. Aerospace Electron. Syst.* **26**(1), 2–11 (1990).
13. B. Bhanu and S. Lee, *Genetic Learning for Adaptive Image Segmentation*. Kluwer Academic Publishers, Boston, Massachusetts (1994).
14. J. Peng and B. Bhanu, Closed-loop object recognition using reinforcement learning, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 538–543. San Francisco, California (June 1996).
15. B. Bhanu, X. Bao and J. Peng, Reinforcement learning integrated image segmentation and object recognition, *Proc. DARPA Image Understanding Workshop*. pp. 1145–1154. New Orleans, Louisiana (May 1997).
16. L. Gupta and M. D. Srinath, Contour sequence moments for the classification of closed planar shapes, *Pattern Recognition* **20**(3), 267–272 (1987).
17. L. Gupta and M. D. Srinath, Invariant planar shape recognition using dynamic alignment, *Pattern Recognition* **21**(3), 235–239 (1988).

18. S. A. Dudani, K. J. Breeding and R. B. McGhee, Aircraft identification by moment invariants, *IEEE Trans. Comput.* **C-26**(1), 39–46 (1977).
19. A. P. Reeves, R. J. Prokop, S. E. Andrews and F. P. Kuhl, Three-dimensional shape analysis using moments and Fourier descriptors, *IEEE Trans. Pattern. Anal. Mach. Intell.* **PAMI-10**(6), 937–943 (1988).
20. G. Y. Tang and T. S. Huang, Using the creation machine to locate airplanes on aerial photos, *Pattern Recognition*, **12**, 431–442 (1980).
21. L. S. Davis and T. C. Henderson, Hierarchical constraint processes for shape analysis, *IEEE Trans. Pattern. Anal. Mach. Intell.* **PAMI-3**(3), 265–277 (1981).
22. C. C. Lin and R. Chellappa, Classification of partial 2-D shapes using Fourier descriptors, *IEEE Trans. Pattern. Anal. Mach. Intell.* **PAMI-9**(5), 686–690 (1987).
23. T. P. Wallace and P. A. Wintz, An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors, *Comput. Graphics Image Process.* **13**, 99–126 (1980).
24. T. P. Wallace, O. P. Mitchell and K. Fukunaga, Three-dimensional shape analysis using local shape descriptors, *IEEE Trans. Pattern. Anal. Mach. Intell.* **PAMI-3**, 310–323 (1981).
25. A. Gorin, Aspect-based aircraft classification from dynamic imagery, *Proc. IEEE Conf. Pattern Recognition and Image Process.* pp. 141–143 (1982).
26. J. W. Gorman, O. R. Mitchell, F. P. Kuhl, Partial shape recognition using dynamic programming, *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(2), 257–266 (1988).
27. Z. Chen and S.-Y. Ho, Computer vision for robust 3d aircraft recognition with fast library search, *Pattern Recognition* **24**(5), 375–390 (1991).
28. J. C. Ming and B. Bhanu, A multistrategy machine learning approach for target model recognition, acquisition, and refinement, *Proc. DARPA Image Understanding Workshop*, pp. 742–756. Pittsburgh, Pennsylvania (September 1990).
29. D. I. Moldovan and C. Wu, A hierarchical knowledge based system for airplane classification, *IEEE Trans. Software Engng*, **14**(12), 1829–1834 (1988).
30. J. Ma, C. Wu, and X. Lu, A fast shape descriptor, *Comput. Vis. Graph. Image Process.* **34**, 282–291 (1986).
31. G. Gibbon, Weighted chord functions, *Pattern Recognition*, **21**(4), 367–379 (1988).
32. J. Ben-Arie and A. Z. Meiri, 3D object recognition by optimal matching search of multinary graphs, *Comput. Vis. Graphics Image Process.* **37**, 345–361 (1987).
33. D. W. Thompson and J. L. Mundy, Three-dimensional model matching from an unconstrained viewpoint, *IEEE Int. Conf. Robotics Autom.*, pp. 208–220. Raleigh, North Carolina (1987).
34. C. H. Chien and J. K. Aggarwal, Model construction and shape recognition from occluding contours. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-11**(4), 372–389 (1989).
35. A. P. Reeves and R. W. Taylor, Identification of three-dimensional object using range information, *IEEE Trans. Pattern. Anal. Mach. Intell.* **11**(4), 403–410 (1989).

About the Author—SUBHODEV DAS received his B. Tech. (Hons.) degree in Electronics and Electrical Engineering from the Indian Institute of Technology, Kharagpur, in 1984, his M. S. degree in Electrical Engineering from the University of Hawaii in 1986, and his Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1991. Between 1991 and 1994, he was with the College of Engineering at the University of California, Riverside. Currently, he is a research staff member at PVI, Inc., Princeton, NJ, working on real-time vision systems. His research interests include intelligent systems, computer vision, machine learning, human-computer interaction, real-time computing, parallel and distributed processing, and applications of artificial intelligence, signal and image processing. He has one patent and about 30 reviewed publications in these areas. He is a member of the IEEE.

About the Author—BIR BHANU received the S. M. and E. E. degrees from Massachusetts Institute of Technology; the Ph.D. degree from the Image Processing Institute, University of Southern California and the M. B. A. degree from the University of California, Irvine. Since 1991, Dr. Bhanu has been a Professor and Director of Visualization and Intelligent Systems Laboratory at the University of California, Riverside. Prior to that he was a Senior Honeywell Fellow at Honeywell Systems & Research Center in Minneapolis. He has been the principal investigator of various programs from DARPA, AFOSR, ARO, NASA, NSF and other agencies and industries. He has 10 patents and over 150 reviewed publications in the areas of computer vision, image processing, pattern recognition, artificial intelligence and learning. He was the Chair for DARPA Image Understanding Workshop, 1994, and the General Chair for the IEEE Conference on Computer Vision and Pattern Recognition, 1996. He is a Fellow of IEEE and a member of Pattern Recognition Society.