Qualitative Recognition of Aircraft in Perspective Aerial Images

S. Das, B. Bhanu, X. Wu, and R.N. Braithwaite

Abstract. Recognition of aircraft in complex, perspective aerial imagery is difficult because of occlusion, shadow, cloud cover, haze, seasonal variations, clutter and various forms of image degradation. This chapter describes a system for aircraft recognition that addresses some of these issues. The recognition system uses a hierarchical object model database that includes models represented using advance concepts to geometric entities. It involves three key processes: (a) The qualitative object recognition process is responsible for model-based symbolic feature extraction and generic object recognition; (b) The refocused matching and evaluation process accesses deeper levels of the database hierarchy with input from (a) to refine the extracted features and to perform more specific classification; 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.

1 Introduction

Photointerpretation (PI) has been an important application domain of image understanding (IU) techniques for about two decades. There have been several U.S. Government programs initiated in the past, e.g., SCORPIUS [8], RADIUS [16, 36], in this area. The importance of automating the PI process is underscored by the fact that a very small percentage of the available imagery in the intelligence community is currently being analyzed due to manpower limitations [1]. An important goal of PI is image exploitation or extraction of intelligence from image data, particularly aerial imagery, to aid reconnaissance tasks, such as airfield, port, and troop movement monitoring. The problem of PI is one of identifying instances of “known” object models in images acquired from a platform, such as by a satellite or a reconnaissance aircraft. The “known” objects are a collection of geometric (e.g., CAD-based) models available to the object recognition system. Model-based object recognition is a challenging task and requires efficient indexing of the stored object models for realistic-time performance.

In real-world PI scenarios, there are additional factors that complicate the overall model-based object recognition process. These include occlusion, shadow, cloud cover, haze, seasonal variations, clutter, and various other forms of image degradation. All of these problems put heavy requirements on any IU system to be robust. For example, the goal of the SCORPIUS program was to be demonstrate the object-oriented image exploitation capability over a wide range of imaging condi-
tions with performance equivalent to a "novice" image analyst [8]. An IU system for photointerpretation is typically required to identify buildings, aircraft, ships, ground vehicles, bridges, and storage facilities. Of these, the first two object classes - 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, such as those mentioned earlier in this paragraph, or are demonstrated to be effective in practical scenarios.

In this chapter, we describe an IU system for aircraft recognition under development 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. We distinguish between two types of image features: primitive or low-level, features that are completely data-driven, e.g., edge segments, lines, regions; symbolic or intermediate-level, features that are data/goal driven, e.g., convex groups of lines, surfaces, ribs. Symbolic features are usually derived from mappings of primitive features under geometrical and physical constraints. In addition to features, we shall also refer to concepts that are used to represent the system's knowledge about the domain of aircraft recognition and about the world. There are two types of concepts: basic, concepts that are independent of the domain, such as unary properties (e.g., small, large), binary properties (e.g., smaller, larger), spatial relationships (e.g., connected-to, left-of, front); advance, concepts that are domain dependent and are defined in terms of symbolic features, basic concepts, and other advance concepts, such as an aircraft wing which can be represented by a convex group of lines and which is known to be connected to the aircraft fuselage.

Our system uses a hierarchical representation, consisting of qualitative-to-quantitative descriptions, of object models (aircraft in this case). Such descriptions 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 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. Since the space for indices may be bounded [12], such distribution of information among the various levels enables one to derive a set of efficient indices for entities at each level. Besides, this representation allows partial object recognition, such as determination of an object class, even when a precise object model is not available. Current model-based recognition systems do not have this capability, i.e., they do not exhibit graceful degradation when encountering a new model.

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 that of camera/
platform, sun, shadow to derive these symbolic features 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, is used to demonstrate the effectiveness of our proposed approach. In the following sections, we will discuss the details of our approach to aircraft recognition. Section 2 describes the background and motivation behind the work reported in this paper. Section 3 describes the key features of our qualitative-to-quantitative approach to object recognition. Section 4 presents the details of an algorithm that integrates feature refinement and object classification. Section 5 gives the details of implementation and the experimental results for qualitative recognition of aircraft. Section 6 presents concluding remarks.

2 Background and Motivation

2.1 Background

There is a variety of object recognition and classification techniques that can possibly be applied to the domain of aircraft recognition. However, the algorithms reviewed in this section are those that have been applied to the specific problem of aircraft identification in both 2-D and 3-D.

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

- **Moment Invariant Techniques** – These techniques use moment invariant features [26] 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 moment 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 [51] 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 that are invariant with respect to starting point, rotation, translation, and scale. The advantage of FD’s 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 FD’s 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 that do not exactly belong to any one of the categories described above.

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

The “knowledge-free,” global techniques are inadequate in the real-world recognition context, since these almost always treat the object of interest in isolation from the rest of the image, i.e., they assume a perfect segmentation of the
<table>
<thead>
<tr>
<th>Techniques</th>
<th>References</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment Invariant</td>
<td>Gupta and Srinath [22]</td>
<td>2D; vector of a sequence of moment invariant functions computed from contour</td>
</tr>
<tr>
<td></td>
<td>Gupta and Srinath [23]</td>
<td>2D; moment function based on distances between each contour pixel and object centroid</td>
</tr>
<tr>
<td></td>
<td>Dudani et al. [15]</td>
<td>3D; invariant features computed from silhouette and its border for different views of each aircraft; Bayes and distance-weighted k-NN classification</td>
</tr>
<tr>
<td></td>
<td>Reeves et al. [40]</td>
<td>3D; normalized moment invariants that are less sensitive to noise than conventional invariants; performance comparable to FD's with/without noise</td>
</tr>
<tr>
<td>Syntactic/Semantic Grammar</td>
<td>Tang and Huang [47]</td>
<td>2D; large, useful structures obtained by removing redundant terminal symbols; localization without classification</td>
</tr>
<tr>
<td></td>
<td>Davis and Henderson [14]</td>
<td>2D; hierarchical approach using only grammatically correct fragment at all levels with constraint propagation; classification into general categories</td>
</tr>
<tr>
<td>Fourier Descriptor</td>
<td>Lin and Chellappa [30]</td>
<td>2D; estimation of FD for the complete contour from partial data</td>
</tr>
<tr>
<td></td>
<td>Wallace and Wintz [50]</td>
<td>3D; normalized FD's for local shape descriptions</td>
</tr>
<tr>
<td></td>
<td>Wallace et al. [49]</td>
<td>3D; FD's for local shape descriptions</td>
</tr>
<tr>
<td></td>
<td>Gorin [19]</td>
<td>3D; combines individual classification results from multiple frames to refine accuracy over time</td>
</tr>
<tr>
<td></td>
<td>Gorman et al. [20]</td>
<td>3D; partial recognition of occluded or overlapping objects using FD's of local features</td>
</tr>
<tr>
<td></td>
<td>Chen and Ho [10]</td>
<td>3D; elliptic FD's that are less sensitive to contour perturbation than regular FD's; reduced set of near neighbors in NN classification for speed</td>
</tr>
<tr>
<td>Model and Knowledge-Based</td>
<td>Ming and Bhanu [34]</td>
<td>2D; Explanation-Based Learning for model acquisition and refinement;</td>
</tr>
<tr>
<td>Techniques</td>
<td>References</td>
<td>Descriptions</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Brooks [9]</td>
<td>Conceptual Clustering for classification 3D; 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</td>
<td></td>
</tr>
<tr>
<td>Moldovan and Wu [35]</td>
<td>3D; 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</td>
<td></td>
</tr>
<tr>
<td>COBIUS [7]</td>
<td>3D; hierarchical representation of objects and constrains; dynamic selection of region or edge segmentation for initial interpretation, followed by model-based resegmentation to extract expected objects</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Ma et al. [33]</td>
<td>2D; normalized shape descriptors faster than FD's and requiring less storage</td>
</tr>
<tr>
<td></td>
<td>Gibbon [18]</td>
<td>2D; weighted chord functions to represent angle chords of an object</td>
</tr>
<tr>
<td></td>
<td>Ben-Arie and Meiri [2]</td>
<td>3D; matching of n-ary relational graphs</td>
</tr>
<tr>
<td></td>
<td>Thompson and Mundy [48]</td>
<td>3D; vertex-pairs as invariant features</td>
</tr>
<tr>
<td></td>
<td>Chien and Aggarwal [11]</td>
<td>3D; quadtree and octree-based object representation; identification using occluding contours</td>
</tr>
<tr>
<td></td>
<td>Reeves and Taylor [41]</td>
<td>3D; recognition based on contour, silhouette, and range imagery</td>
</tr>
</tbody>
</table>

object, which is generally not achieved. Among the past research in aircraft recognition, the ACRONYM system by Brooks [9] is closest to the system described in this paper. In ACRONYM, symbols called quantifiers act as place holders for variable quantities in the representation of generic objects. These quantifiers are provided restrictive set of constraints to model more specific subclasses. The
constraints are usually ranges of numbers that determine the degree of specificity of the subclass models. Variations of size, structure, and spatial relations within object classes allow determination of observational invariants of objects that are useful for recognition. Additionally, rules aid in reasoning about geometric relationships between coordinate systems linked by multiple partially specified coordinate transforms. The combination of these two mechanisms provide ACRONYM the means for predicting the appearance of objects in the images. The interpretation process involves the merging of hypothesized local matches (of objects to image features) that provide a consistent global interpretation about the hypothesized object. Using the initial set of derived image primitives, the system searches small image regions for particular features for finer classification and pose estimation of objects.

In spite of its representational elegance, the ACRONYM system falls short of addressing the real-world concerns posed by the PI problem. In particular, it has no mechanism for automatically acquiring and refining object models, and shadows, clutter and other image degradations are not adequately handled within its framework. In other words, this system is not applicable to real-world, complex images without significant enhancements and modifications.

The deficiencies of the current aircraft recognition techniques emphasize the need for a more practically oriented approach, the motivation for which is now presented in the following subsection.

2.2 Motivation

Central to the model-based approach to recognition lies the implicit assumption that each instance of an object is an accurate projection of the object of known dimensions and shape onto the image plane. Consequently, the focus of the model-based work has been to recover that single object from the model database subject to the condition that its descriptions agree best with the projected data for that unknown viewpoint.

The one-step recognition of single-model instances is bound for high complexity and inefficiency because of the lack of reliable low-level image analysis operations and tools. For example, in spite of the large number of image segmentation techniques presently available [25], no general methods have been found that perform adequately across a diverse set of imagery. It is well known that some form of segmentation of the raw image primitives into meaningful groups is essential for any model-based approach [21], whose complexity is exponential in the number of features and linear in the number of models used in recognition. In real-world, complex images, such as those encountered in the PI context, a good segmentation cannot be guaranteed in general. This difficulty is compounded by the fact that a specific aircraft type (e.g., a C-130) can have very different appearances depending on the imaging conditions. To reduce the search time during model matching and also to minimize the false alarm rate, an efficient indexing
scheme needs to be devised. Such an indexing scheme still remains an important but elusive goal of model-based object recognition. Since one has little control on the imaging conditions and hence on the types of features that may be obtained in unstructured environments, a more pragmatic approach to reduce the burden on the indexing scheme would be to adopt a multi-step recognition approach that proceeds from identifying generic object classes to specific object instances. A nodal 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. The conceptualization at any given node is responsible for aggregating the more generic (global/coarse/qualitative) object features spanning the lower nodes while leaving the more specific (local/fine/quantitative) features to the latter. Interestingly, the hierarchically structured model database and the nature of the associated recognition process are so much akin to the human behavior unlike the traditional model-based approaches. Categorization or the process of treating nonidentical stimuli as equivalent in the case of humans is performed in a highly deterministic way and very often occurs at a basic level [42]. Basic categories are those that carry most distinguishing features, such as aircraft, bridge, house, tree, and typically are the ones that are first discriminated in an environment and therefore are more efficiently indexed [42].

Recognition of objects in real-world scenarios is a difficult problem even if the best indexing scheme is available. The difficulty is caused by image variations due to noise, shadow, occlusion, weather condition, etc. Most geometric model-based object recognition approaches rely on spatially local image features, such as edges, for making coarse-level decisions that interact to arrive at a globally consistent interpretation of the observed scene. The aforementioned image variabilities cause missing, spurious, or ambiguous features that contribute to inefficient search of the decision space. Quite often a significant amount of contextual or other scene-specific information is required to alleviate this problem.

One of the key factors for poor recognition performance is the presence of shadows. Shadows may cause occasional loss in photometric information. However, shadows provide valuable 3D information, such as surface orientations [45], distance between surfaces in an image [17, 27, 32]. They have been successfully utilized in recognizing buildings in aerial photographs [28, 29], but have found little attention in aircraft recognition scenarios. One of the ways shadows can be helpful is in the identification of image features that are caused by raised structures, such as buildings, poles, or aircraft, and also in the determination of their 3D 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 factor which a robust IU system needs to address. In the scenario of aircraft recognition in aerial images, occlusion is primarily due to self-occlusion or close parking of aircraft.

There are significant merits in intelligent feature selection for any IU task. Feature selection is of prime importance in executing complex tasks like object recognition in unstructured environments. For the aircraft scenario, this implies an ability to distinguish between image features that are caused by the aircraft

3 A Framer
structure and those that constitute the clutter. An early selection of the relevant image features could significantly reduce the burden on the subsequent interpretation stage, which would otherwise need to distinguish among several alternatives hypothesized by the different groups of local features. One way to determine the relevant features is to identify the perceptually salient image contours (linked edge segments). As psychological evidences indicate, very often such contours are associated with objects that could be of interest or that 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.

Further improvement in aircraft image understanding is possible by integrating the feature selection and image interpretation processes, so that the interpretation process is driven by the most salient contours first, followed by the less salient ones. Previous attempts at aircraft recognition have focused on a single type of image features, 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 [7]. 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 in meaningful ways.

3 A Framework for Model-Based Object Recognition

The general problem of model-based object recognition encompasses the following issues: representation of object models, prediction of observables (mapping of models to images), selection of image features, matching of selected features to the predictions (mapping of images to models), and resolving inconsistencies during the matching process in order to deduce a globally consistent 3D interpretation of the image. We now describe our framework for geometric, model-based object recognition which addresses these issues. Subsequently, the framework forms the basis of our aircraft recognition system. The framework, which is schematically shown in Fig. 1, has four key features: (a) a qualitative-to-quantitative hierarchical object model database that provides the object models for the domain of recognition, and three recognition sub-processes that utilize these models – (b) saliency-based regulation of low-level features that extracts these features from the input images and makes them available to the subsequent steps of recognition in an incremental fashion, (c) model-based symbolic feature extraction and evaluation step that uses the regulated low-level features and the object models together with heterogeneous models of image segmentation, shadow casting, and image acquisition to derive symbolic features for generic object recognition, and (d) refocused matching and evaluation step to perform finer object classification using the hierarchical model database. These four features are now described in detail.
3.1 Qualitative-to-Quantitative Hierarchical Object Model Database

Each level of the representation hierarchy indicates a certain degree of conceptualization for the object entities. The conceptualization of objects at any given level has "class"-specific features sufficient enough to resolve classes at that level. The choice of generic object representation can be in the form of structural models [3, 6, 9] or it can be based on the functional aspect of the objects [46]. In our approach, we have adopted the former since it is the most natural way to express the part-whole relationships in an object model and also the conceptualization is based on the structural components of the model. However, it is almost intuitive that the best object recognition system for a given task should combine both of these representations.

The top-level of the hierarchy corresponds to the basic categories [42], e.g., chair, aircraft. According to Rosch et al. [42], the members of each basic category share most common attributes which may be used to form a definite image of that category in human memory. In contrast, the categories that are one level more abstract or the superordinate categories (e.g., furniture, vehicle) exhibit little commonality among its members, while those below the basic level or the subordinate categories (e.g., lounge chair, jumbo jet) demonstrate significant overlap of attributes with other similar categories (for example, jumbo jet shares most of its attributes with other kinds of jets). Thus, the basic categories in the object recognition context are those which can be described in the most conceptual form using image-based structural information. The description of a basic category includes its shape attributes and structural subparts in a symbolic (qualitative) form. The progressively deeper levels embody more specific knowledge that becomes completely quantitative once the terminal nodes (location of geometric models) have been reached. A partial hierarchy illustrating this particular database structure for the aircraft category is shown in Fig. 2.

There are two important considerations in designing such a hierarchical database from the point of view object recognition:
• the choice of features to represent a particular object class. In our framework, it is driven by the discriminating power of the features in distinguishing among objects at the same level of the hierarchy. Besides, these 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 to support that object hypothesis.

• the matching process. In our framework, the process can search a lower level for distinguishing features should a categorization be not possible at a particular level because of the lack of suitable features derived from the image (a realistic situation). Thus, the flow of control during matching is bi-directional, between a generalized class and its more specialized subclasses.

The ACRONYM [9] system which also adopts a hierarchical representational scheme of generic and specific objects addresses these two issues in somewhat different ways. In ACRONYM, the representation scheme is based on the specialization (of an object class) primitive. In other words, all the features that 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. In ACRONYM, the reasoning about object classes and their specializations is carried out at the same time, i.e., matches to some parts of an object class is automatically carried down to a match to the corresponding part of a specialization of that object class for the purpose of mensuration. 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.

![Generic-Aircraft Diagram](image)

**Fig. 2.** A partial hierarchy of generic-to-specific aircraft
To specify a subclass of a selected object class, new structural features that are essential for subclass discrimination may be introduced in a qualitative manner or known features (belonging to the parent class) may be instantiated as in ACRONYM. For example, 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 as shown in Fig. 2. Another important issue related to this hierarchical design is the degree of conceptualization that may be required at each level which in turn determines the number of levels. However, that is not the focus of this paper, instead the emphasis here is on given such a hierarchical database of advance concepts to geometric entities how one may utilize it to perform qualitative-to-quantitative recognition of objects.

3.2 Saliency-Based Regulation of Low-Level Features

Application of current model-based object recognition techniques to real-world, complex scenarios generally produces unsatisfactory results because of the large amount of image clutter that may be present in these cases. The clutter information misleads the recognition system to believe that it is associated with the target objects when actually it is not. A preprocessing step which is so important for any IU task has been ignored, by and large, in model-based object recognition approaches. At the same time, considerable effort has been directed towards segmentation and grouping of primitive features. The preprocessing step is necessary to improve object contrast and to reduce noise and clutter in the image. While noise and contrast may be improved by applying standard filtering techniques, clutter rejections 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. 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. Sha'ashua and Ullman [44] present an iterative method of extracting a collection of data that are globally salient using local measurements of curvature and curvature variation.

We adopt a non-iterative approach that 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 that may be encountered in practice, let alone for the different parts of a single image. Edge pixels are grouped based on the local curvature values and the continuity of edge segments. In addition, the flow of low-level features that are used to derive the symbolic features is regulated based on the edge-strength measure. Regulation may also be based on the "specialized" nature of the features as required by the refocused matching process.

3.3 Model-Based

This model is basic level. Because it is crucial that the matching groups of primitive features have the discriminative power and should have the discriminatory power.

The main drawback of the symbolic features is category. The evidence on the domain is that the symmetric production rule for the action asserts that there are two types of extraction:

- a qualitative and advance
- a quantitative features or attributes

For example, input the part is up the action is the rule. Such whole [3] and is centered.

The extraction of the special considers that the model is successful, the problem hierarchy and is visited in a goal-directed perceptual processing is started. In most cases what matters is to reduce the number of feature extraction.

The evaluation components to be...
3.3 Model-Based Symbolic Feature Extraction and Evaluation

This module is responsible for generic object recognition or categorization at the basic level. Because of the conceptual nature of the model descriptions at this level, it is crucial that symbolic features be extracted from the primitives and be used in the matching process. Symbolic features may be derived using perceptually grouped primitives. The attractive feature of perceptual grouping is that the groups are more useful for higher-level image interpretation processes, such as object recognition, than the individual primitive features of the group and the grouping process is domain independent. Although, the perceptually grouped primitives improve the efficiency of recognition search techniques, they do not have the discriminating power of symbolic features that are domain dependent. The main drawback of the perceptual groups is that these are based on local measures, such as proximity, collinearity, parallelism [31], while the desired symbolic features need to exhibit more global properties that are typical of a basic category. The extraction of symbolic features is goal-driven in our approach, based on the domain knowledge about generic objects.

Both symbolic features and advance concepts are represented in our system by production rules. Once the conditions of a rule have been satisfied, the rule action asserts the presence of the symbolic feature or the advance concept. There are two types of production rules:

- a qualitative rule, which measures the qualitative properties of symbolic features and advance concepts, and
- a quantitative rule, which computes values for the corresponding symbolic features or advance concepts.

For example, in the case of a base category, the conditions of the rule are essentially the part decompositions and the structural relationships among the parts, while the action is the name of the category. Each part is defined by a similar production rule. Such whole-to-part decomposition is essential for generic object recognition [3] and is central to the Recognition By Components (RBC) theory [6].

The extraction of symbolic features beyond perceptual grouping requires special consideration. This is because there exists no “meta-rule” of perceptual organization that will allow combination of its domain-independent rules. Consequently, the recognition search process is initiated at the top level of the database hierarchy and is allowed to loop through the production rules of the node being visited in a goal-decomposition fashion till it encounters a rule whose conditions involve perceptual groups. At this instant, the symbolic feature extraction process is started. In most practical situations, the need for recognition will be driven by what matters in the environment. Therefore, the perceiving agent will be able to reduce the number of base categories that need to be visited to initiate the symbolic feature extraction process by using some background/contextual knowledge.

The evaluation process involves the verification of the global semantic shape components to correspond to the qualitative object or the base category, i.e., the
verification step is one of testing all the conditions of the highest level production rule associated with a base category. It is essentially a reasoning process based on evidence accumulation to infer the presence of the selected base category in the image. Such decision making process is typical of humans and constitutes the very last stage in the feature analysis model of human pattern recognition [43]. The key feature of this evaluation step in our framework is the use of heterogeneous models to accrue evidence for the semantic components of a base category. These models include:

- edge/grey scale-based model for image segmentation,
- models for shadow casting process, and
- models for image acquisition.

Also used are the dominant axes that characterize the shape of the generic object class. We assert that the use of such heterogeneous models is essential for any real-world aircraft recognition task and is exclusive to our recognition system.

Finally, our system provides a feedback path from the generic recognition module to a feature regulation module (see Fig. 1). The feedback control is used 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. The output of the symbolic feature extraction and evaluation process is a known generic object class and labeled symbolic features that will be useful for finer classification.

3.4 Refocused Matching and Evaluation

This module is responsible for further classification of an object whose category has been determined by the generic recognition process. The key difference of this step and the multiresolution approaches to recognition is that the former views the object model database at increasing resolution instead of the image. The additional effort involved in our approach is the creation of the hierarchical database, while in the latter the resolution is varied by image resampling. 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. In our approach, the data reduction is achieved by deriving symbolic features that are more “focused” or localized with respect to a particular level.

During the course of refocused recognition, instances of new symbolic features of the object model may be identified in the image or old symbolic features extracted in previous cycles may be subjected to mensuration. Evaluation of the recognition results may require refinement of the features at the symbolic feature extraction level or new model features may prompt access to “special” local features (e.g., a curve at a certain image location) at the feature regulator level.
Usually, the derived symbolic features guide the search for these special features. The final result of refocused matching is object recognition/classification.

## 4 A Qualitative Recognition Algorithm

The emphasis of this paper is on qualitative recognition of aircraft using the framework described in the previous section. In our representational scheme, the volumetric shape of a generic aircraft is described using the linear RSHGC (Right Straight Homogeneous Generalized Cylinder, see Fig. 3) representation [4] and the structural components of a generic aircraft are wings, fuselage, tails, rudder, and nose (see Table 2). The corresponding (expected) image descriptions are indicated in Table 3. These predictions about image features follow the fact that linear RSHGC descriptors project to lines in a plane (see Appendix A.1). The symbolic definitions (simplified) of a generic aircraft and its three subclasses based on the hierarchy of Fig. 2 are illustrated in Fig. 4. Here, the action part is the assertion that the symbolic description is one of a generic aircraft or one of its subclasses. Similar production rules are used to define each of the subparts. One common feature of the shape descriptions of all the subparts is the convexity and another one is the single axis of symmetry. Examples of subpart-specific information include trapezoid-like shapes, one that monotonically decrease away from the fuselage, for the wings and tails, and the quantification of these subparts (see Table 3). At the subclass level, the distinctions are based on the engine location - on-wing (large), on-fuselage (large/medium), in-body (small) - which is still a qualitative information but is more specific. All of these symbolic definitions are stored in the hierarchical object model database described in Sect. 3.1.

Give an input 2-D image and ancillary data about the imaging parameters and scene conditions, the algorithm for qualitative recognition of aircraft consists of

Fig. 3. Generalized Cylinders representation of the subparts of a generic aircraft: a cross-sections of GCs, and b axes of GCs in a
Table 2. The description of the form of a generic aircraft model

<table>
<thead>
<tr>
<th>Model Part</th>
<th>Qualitative Definition</th>
<th>Inter-relationship</th>
</tr>
</thead>
</table>
| Nose       | • one point of high curvature (nose tip)  
                         • one axis of symmetry through this high curvature point 
                         • monotonically increasing cross-section along this axis and at right angles to it starting at one end | connected to the fuselage |
| Fuselage   | • elongated shape 
                         • one axis of symmetry 
                         • monotonically increasing right cross-section along this axis | connected to the nose, wings, tails, rudder |
| Wing       | • one axis of symmetry 
                         • monotonically increasing right cross-section along this axis 
                         • two in number placed equidistant from the nose tip on either side of the fuselage axis | connected to the fuselage-the axis of symmetry of each wing makes the same angle with the fuselage axis |
| Tail       | • one axis of symmetry 
                         • monotonically increasing right cross-section along this axis 
                         • two in number placed equidistant from the nose tip on either side of the fuselage axis | connected to the fuselage-the axis of symmetry of each tail makes the same angle with the fuselage axis |
| Rudder     | • one axis of symmetry 
                         • monotonically increasing right cross-section along this axis | connected to the fuselage-the axis of symmetry is in a plane perpendicular to that of the fuselage axis |

Table 3. Image descriptions of the different parts of the generic aircraft model

<table>
<thead>
<tr>
<th>Model part</th>
<th>Model shape</th>
<th>Image contour shape</th>
<th>Connecting (to other parts) segment shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nose</td>
<td>Straight</td>
<td>Wedge</td>
<td>Straight</td>
</tr>
<tr>
<td>Fuselage</td>
<td>Straight</td>
<td>Linear</td>
<td>Straight</td>
</tr>
<tr>
<td>Wing</td>
<td>Straight</td>
<td>Trapezoid</td>
<td>Straight</td>
</tr>
<tr>
<td>Tail</td>
<td>Straight</td>
<td>Trapezoid</td>
<td>Straight</td>
</tr>
<tr>
<td>Rudder</td>
<td>Straight</td>
<td>Trapezoid</td>
<td>Straight</td>
</tr>
</tbody>
</table>

the steps illustrated in Fig. 5. Some of these steps, i.e., the critical ones, such as saliency-based features regulation, symbolic feature extraction and generic object recognition, and refocused matching based on specialized feature acquisition, have been discussed in Sect. 3 while describing our framework for model-based
Qualitative Recognition of Aircraft in Perspective Aerial Images

\[\text{(define-rule GENERIC-AIRCRAFT} \]
\[\text{ "The description of a generic aircraft"} \]
\[\text{(shape-description = Linear RSHGC)} \]
\[\text{(symbolic-feature = WING) (satisfy = TRUE)} \]
\[\text{(symbolic-feature = FIN) (satisfy = TRUE)} \]
\[\text{(symbolic-feature = TAIL) (satisfy = 90)} \]
\[\text{(symbolic-feature = NOSE) (satisfy = 90)} \]
\[\text{(connected-to (WING, FUSelage))} \]
\[\text{(connected-to (TAIL, FUSelage))} \]
\[\text{(connected-to (NOSE, FUSelage))} \]
\[\text{(closer-to (WING, NOSE, TAIL))} \]
\[\text{(closer-to (WING, NOSE, RUDDER))} \]
\[\text{(closer-to (TAIL, RUDDER, WING))} \]
\[\text{(closer-to (RUDDER, TAIL, WING))} \]
\[\text{(define-rule LARGE-AIRCRAFT} \]
\[\text{ "The description of a large aircraft class"} \]
\[\text{(symbolic-feature = ENGINE) (satisfy = 90)} \]
\[\text{(location (ENGINE, WING))} \]
\[\text{(define-rule MID-AIRCRAFT} \]
\[\text{ "The description of a mid aircraft class"} \]
\[\text{(symbolic-feature = ENGINE) (satisfy = 90)} \]
\[\text{(location (ENGINE, WING))} \]
\[\text{(define-rule SMALL-AIRCRAFT} \]
\[\text{ "The description of a small aircraft class"} \]
\[\text{(symbolic-feature = ENGINE) (satisfy = 90)} \]
\[\text{(location (ENGINE, WING))} \]

Fig. 4. Simplified examples of advance concepts: a a generic aircraft, and the three aircraft classes; b large; c medium; d small

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

Object recognition. These steps, along with others shown in Fig. 5, will now be detailed as part of the qualitative recognition algorithm.

4.1 Region of Interest Identification

In any 3D scene interpretation scenario, the objects of interest typically occupy a limited portion of the visual field. On the other hand, the input imagery in PI are
usually of very high resolution, which puts significant burden on preprocessing, segmentation, and feature extraction steps. By performing intelligent search to direct the focus of attention at certain parts of the image, the regions of interest (ROIs) containing the target objects (aircraft in our case) can be identified. A traditional approach is to sample the image and submit the low resolution image to the recognition process, but such subsampling may lose important information regarding the object of interest. Alternately, we adopt a context-driven approach to isolate the basic categories or generic classes from the database that are likely to be present in the image. Our approach begins with a low resolution version of the original image and identifies different image parts that have a large concentration of perceptual groups of primitive feature types that are characteristic of the targeted generic classes. Next, each of these subimages are examined at increasingly higher resolution using similar argument until the original resolution is attained. Further details of this step appears in [38].

Once the ROIs have been identified, each of them is subjected to the recognition process in succession. The following discussions apply to every ROI.

### 4.2 Saliency-Based Feature Regulation

An ROI is first treated with a Gaussian filter for noise removal. Edge pixels in images are detected using six $5 \times 5$ masks sensitive to edge orientations in steps of $30^\circ$ [39]. At each image location, the edge response is set to the magnitude (and the corresponding direction) of the $5 \times 5$ edge mask that gives the largest response. Edges are thinned to one-pixel width by retaining only those edge pixels that satisfy the following conditions: (a) the edge magnitude is larger than the magnitude of its two neighbors in a direction normal to the edge direction, and (b) the edge directions of the two neighboring pixels are within $30^\circ$ of that of the central pixel. The thinned edge pixels are next used to extract perceptually salient contours and other primitive features. Intuitively speaking, the strongest edge pixels will typically lie on the object boundaries. However, what separates an object from the background in monochromatic images are the perceptually salient edge contours or linked edge segments. Our approach to identifying perceptually salient contours is based on finding long, smooth edge segments that are made up of high-magnitude edge pixels.

The above goal can be 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), subject to the following optimization:

$$\max_{c \in \mathcal{C}} \left[ \int_c [\omega_x \Delta(s) + \omega_y \lambda(s)] ds - \omega_x \int_c (d\theta(ds)^2) ds \right].$$

(1)

Here, $\mathcal{C}_N$ denotes the set of all contours, $C$, of length $N$ beginning at $s = 0$. The variable $\Delta(s)$ is the magnitude of an edge pixel along the contour and denotes the...
strength component of the criterion function; \( \lambda(s) = 1 \) if \( \Delta(s) \) is greater than a chosen threshold and \( \lambda(s) = 0 \), otherwise, and it represents the length component; the variable \( d\theta/ds \) 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. The constants \( \omega_1 \), \( \omega_2 \), and \( \omega_3 \) are the weights of the strength, length, and smoothness components of the criterion function, respectively, where \( 0 < \omega_1, \omega_2, \omega_3 < 1 \). We note that (1) is a non-linear optimization in terms of the variable \( C \), and an exhaustive search would span a space whose size is \( p^N \), \( p \) being the number of pixels that need to be considered at each point along the contour \( C \). We therefore reduce the complexity of the above optimization problem by adopting a multi-step approach. To do this, we decompose (1) into several optimization steps:

\[
\max_{C \in \mathcal{C}} \left[ \int_C \left( \omega_1 \Delta(s) + \omega_2 \lambda(s) \right) ds - \omega_3 \int_s \left( d\theta/ds \right)^2 ds \right] = \max_{C \in \mathcal{C}} \left[ \int_S \left( \omega_1 \Delta(s) + \omega_2 \lambda(s) \right) ds - \omega_3 \int_s \left( d\theta/ds \right)^2 ds \right] + \max_{C \in \mathcal{C}} \left[ \int_S \left( \omega_1 \Delta(s) + \omega_2 \lambda(s) \right) ds - \omega_3 \int_s \left( d\theta/ds \right)^2 ds \right] + \ldots + \max_{C \in \mathcal{C}} \left[ \int_S \left( \omega_1 \Delta(s) + \omega_2 \lambda(s) \right) ds - \omega_3 \int_s \left( d\theta/ds \right)^2 ds \right],
\]

where \( N = \sum_{i=1}^{N} N_i \) and \( N = \sum_{i=1}^{n} N_i \). Now, the multi-step optimization problem is concerned with finding segments, \( S \), that are of lengths \( N_1, \ldots, N_n \), each of which is smaller than the original contour, \( C \), of length \( N \). Further reduction in complexity is achieved if the number of pixels, \( p \), that need to be considered at each point along the segment is lowered. Both of these conditions, viz., shorter edge segments and fewer neighboring pixels, are satisfied by thresholding a thinned edge image using multiple thresholds. The application of multiple edge-magnitude thresholds results in a set of edge images containing segments of edge contours in the original thinned edge image.

To create long chains of edge segments, we initiate edge segment-following at a terminal pixel (one which has a single neighboring edge pixel) in the edge image obtained with the highest threshold value. The edge-segment following process is accomplished by selecting a neighbor of a given edge pixel of the segment that maximizes the optimizing criterion of (1). To account for noise, our approach allows a gap length of two pixels to be covered by the edge segment-following process. Since a very long contour would rarely appear in a single edge image, the edge segment-following process continues across edge images obtained with progressively lower thresholds after it has been terminated within a single edge image. The process stops when the current edge image is the last of the edge image set, only to be restarted at another terminal pixel (belonging to a different contour) in the first image. This new starting pixel is selected through a raster scan and the subsequent repetition of the edge segment-following process yields a different
4.3 Primitive Feature Extraction

In order to extract primitive features for generic object recognition, the feature extraction process is based on the expected image descriptions of the generic classes. According to Table 3, such descriptions for a generic aircraft consist of linear segments. In our system, we have implemented a line extraction algorithm, based upon the detection of significant instances of collinearity (scale-independent) among edge points, similar to the one proposed by Lowe [31]. The input to this line extraction algorithm is a set of regulated edge segments. In addition to lines, our algorithm also detects corners by obtaining gradient and curvature measurements at pixels in the grey scale image [37].

4.4 Region Segmentation and Dominant Axes Extraction

Given the input ROI image, the region segmentation process uses grey 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 [5]. The joint relaxation provides edge and grey value interactions in the initial label (probability) assignment to 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.

Extraction of dominant axes that characterize the shape of a generic object in the image, requires access to the model of the shape. In our algorithm, the potential dominant axes of the generic aircraft shape are generated by connecting the extremities of a labeled region within a segmented image. To determine the extreme points, the smallest convex polygon (in terms of area) that completely surrounds the object region is found. The smallest convex polygon is one whose vertices lie close to the local extrema of curvature points along the boundary of the labeled region. The extremities of the segmented region should correspond to these local extrema of curvature points that are near the vertices of the smallest

4.5 Potential

Shadows are an important feature in general method for object recognition. The shadow is cast upon the object as the light illuminates it from a certain direction. The shadow can be categorized into two types:

- Convex: The shadow is cast from the projection of the object onto a plane.
- Concave: The shadow is cast from the projection of an object onto a plane that is not parallel to the light source.

Assuming that the shadow is represented by straight lines, the shadow consists of two lines that are parallel and equidistant from each other. These lines are the projection of the object onto the plane and the projection of the light source onto the plane. The distance between these lines is equal to the distance between the object and the light source.
polygon. Now, more than one "extreme" point may be identified within a small neighborhood along the region boundary in the vicinity of a polygon vertex, e.g., when there are multiple local extrema points in a certain segment of the boundary. In that case, nearby "extreme" points are grouped into clusters and the cluster centers are chosen to represent the region extremities. A potential dominant axis is a line that connects two such extreme points that are not the centers of adjacent clusters. Since a perfect segmentation is difficult to achieve in practice, the segmented region may not have the exact shape of a generic object or it may be fragmented. Thus, it is additionally ensured that a potential dominant axis is not located outside the segmented region. Lines whose significant portions are not contained within the segmented region are ignored.

4.5 Potential Shadow Identification

Shadows are unavoidable characteristics of outdoor scenes. Some have exploited shadow information in a scene-specific manner [28], while others have suggested general methods of shadow interpretation [24]. The shadow of an object has two parts: the object's dark side (or self-shadow) and its projected (or cast) shadow. The shadow boundary is defined by a brightness discontinuity between directly illuminated regions and regions receiving either no light or only indirectly scattered light. It is important to note that the information about the shadow casting object is contained only in the shadow boundaries. A shadow boundary has three basic types of segments (see Appendix A.2): shadow, shadow-making, and occluding. The first type belongs to the cast shadow, whereas the rest are associated with the shadow-casting object. The following are some useful properties of shadow edges (2-D projections of shadow boundaries) [32]:

- Contrast across a shadow edge is equal to the ratio of direct to indirect light. This ratio along the length of the shadow will change only smoothly and independently from the surface on which it falls. For a distant light source, the ratio is nearly constant.
- Each point on a shadow edge corresponds to some point on the shadow-making edge in the direction toward or away from the illumination vanishing point (projection of the point source).
- Shadow breaks (tangent discontinuities or corners) are caused by breaks in shadow-making edges unless the source of illumination is coincidently coplanar with the two tangents of the break or shadows cast by different objects intersect.

Assuming that any arbitrarily shaped shadow boundary can be locally represented by straight lines, we have developed an algorithm to detect potential shadow lines. It is based on the test of bimodality of the local histogram which is a consequence of the first property of a shadow edge mentioned above. A rectangular window is first set up on each side of a selected line. The dimensions of the window are \( l \times w \), where \( l \) is the length of the line and the width \( w \) is chosen to be
large enough for the histogram computation to be effective. Within each window, the region segmentation algorithm described above is applied and the largest region is retained. Next, a grey-level histogram is obtained for the region and the significant modes are identified.

Identification of modes involves locating the peaks and valleys in the 1-D histogram first, followed by clustering of the grey-levels between two consecutive valleys. Small clusters are removed from consideration. Given that the grey levels in the k-th cluster are between $p_k$ and $q_k$, the mode of the cluster is obtained as

$$m_k = \frac{\sum_{i=p_k}^{q_k} iN(i)}{\sum_{i=p_k}^{q_k} N(i)} \quad (3)$$

where $N(i)$ denotes the frequency of the i-th grey level. For two adjacent clusters, whose modes are $m_k$ and $m_{k+1}$, and whose total number of pixels are $N_k$ and $N_{k+1}$, respectively, the following averaged mode is computed:

$$\hat{m} = \frac{m_k N_k + m_{k+1} N_{k+1}}{N_k + N_{k+1}} \quad (4)$$

where $N_r = N_k + N_{k+1}$. The two clusters are merged if

$$(m_k - \hat{m})^2 N_k / N_r < T_1 \quad \text{and} \quad (m_{k+1} - \hat{m})^2 N_{k+1} / N_r < T_1.$$  

The new mode for the merged clusters is obtained by applying (3). The process is continued until there is no change in the number of clusters. Finally, the mode corresponding to the largest cluster is retained as the significant mode. Since the histogram is based on the pixels of a single region, there is usually a single significant mode. The two most significant modes from either side of the line are then subjected to the bimodality test. If the separation between the modes is less than a threshold, $T_1$, or the smaller of the two is more than another threshold, $T_2$, the line is ignored. Otherwise, it is marked as a potential shadow edge.

### 4.6 Non-Shadow Feature Extraction

The shadow edges detected in the previous stage (Sect. 4.5) may include projections of shadow, shadow-making, and occluding boundaries. The focus of this step is to separate the shadow boundaries from the rest. In order to do this, we make use of the remaining properties of shadow edges stated in the preceding subsection. Additionally, we utilize the ancillary data about the camera-platform position/orientation and the illumination point source, $I_p$ (the sun in our case) position together with the imaging parameters to compute the image plane position of the illumination projection point, $i_r$. The shadow edges are identified by pairing them up with shadow-making edges while observing that the illumina-
Qualitative Recognition of Aircraft in Perspective Aerial Images

nation rays diverge from \( i, (I, \text{ in front of the camera}) \), converge to \( i, (I, \text{ behind the camera}) \), or are parallel \((I, \text{ in the lens plane of the camera}) \). However, the location of \( i \) itself is not a strong constraint to resolve correspondences between shadow and shadow-making boundaries except for specific aspects, e.g., direct overhead-view. The lines therefore need to be grouped in some meaningful way such that two groups would correspond to each other only when their members do.

One way to group the lines is to form their convex sets. A convex group of shadow-making lines would normally cast a convex group of shadow lines. Initially, groups of lines are formed based on the perceptual cues of proximity and collinearity [31]. Each group is then subjected to a convexity test for which we utilize the segmentation results for the entire ROI from Sect. 4.4. The motivation here is that if the elements of the 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.

We recall that the convexity property requires that the line joining any two points on the boundary of a convex figure must be completely contained within that figure. A convexity test is performed for every pair of lines in a selected group. We, however, relax the “containment” condition since the region borders do not precisely coincide with the edge borders. For any two lines \( l_i \) and \( l_j \) in a selected group, we determine the two lines, \( l_{ii}^k \) and \( l_{ji}^k \), obtained by joining the end points \((e_i, e_j)\) of \( l_i \) and \((e_j, e_i)\) of \( l_j \). If \( R \) denotes the corresponding segmented region, then \( l_{ii}^k \) and \( l_{ji}^k \) will be said to be completely contained in \( R \) if

\[
\frac{N_{ii^k}}{N_{ii}} \geq T_s \quad \text{and} \quad \frac{N_{ji^k}}{N_{ji}} \geq T_s.
\]

Here, \( N_s \) denotes the number of pixels in the set \( s \).

If there is a line in a group which fails the above convexity test when paired with any other line from the same group, 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.

Once the convex groups of lines have been identified, for each marked shadow line in a group, a corresponding shadow-making line from another group is sought by searching in a direction towards \((\text{or away from})\ i, \). The matching score is determined by the degree of overlap of the two matching lines in the predicted direction. Let \( r \) be a shadow line from a convex group, \( G_r \), and let \( t \) be another line \((\text{or may not be marked as a shadow})\) belonging to a different convex group, \( G_s \). Also, let \( \hat{n} \) denote the unit vector connecting the \( j \)th pixel of \( r \) and \( i, \) and let \( \hat{n}_s \) denote that connecting the \( j \)th pixel of \( t \) and \( i, \). Then the total matching score of the pair \((r, t)\) is
\[ \phi_r = \sum_{i=1}^{N_r} \sum_{m=1}^{N_m} \frac{\hat{n}_{r_i} \cdot \hat{n}_{m}}{N_r + N_m}, \]

where \( N_r \) and \( N_m \) denote the lengths (in pixels) of the lines \( r \) and \( m \), respectively. In order to minimize the match of one pixel element of one pixel element of \( r \) with multiple elements of \( m \), it is ensured that the angle \( \theta_m = \arccos(\hat{n}_{r_i} \cdot \hat{n}_m) < T_r \).

All the candidate matches of a selected shadow line are arranged according to the matching scores and marked with the corresponding group identifier. For example, the line \( r \) corresponding to the selected shadow line \( r \) is assigned the label \((\phi_r, G_r)\). This entire matching process is repeated for all other shadow lines in that particular group \((G_r, \text{in the above example})\). 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., non-shadow) group.

### 4.7 Symbolic Feature Extraction

Since symbolic features are characteristic of generic classes and are more global than the perceptually grouped features, their extraction tends to be domain-dependent or more specifically goal-driven. According to Table 3, the symbolic features for a generic aircraft class included trapezoid-like shapes for wings, tails, and rudder, and wedge-like shape for the nose part. In Sect. 3, we outlined the steps of our symbolic feature extraction step while describing the framework of our object recognition approach. Here, we additionally note that the perceptual groups that would be used in symbolic feature extraction must come from the non-shadow features. Our algorithm has already identified the convex groups of non-shadow features in the preceding step and convexity is a common characteristic of most geometric-modeled objects. These convex groups are now used to derive the specific types of symbolic features.

To identify the trapezoid-like shapes, groups of three (partially closed contour) and four (fully closed contour) lines are considered. The partially closed contours are typical of any aircraft wing, tail, and rudder sections, while the fully closed contours are a rarity. Any group of three lines must satisfy the following conditions:

- two lines are non-parallel (using parallelism measure of [31]),
- the third line is in between the two non-parallel,
- the intersection of the third line with the non-parallel occur near independently detected corners, and
- the third line is smaller in length than one of the non-parallel (at least).

To overcome the lines into smaller above sets of conditing line pairs that enforcing the cond proportional to the [31] is based on th can therefore be sir an additional perce only. Given the poi endpoints of these measure, \( \text{prox}_{np} \) for

\[ \text{prox}_{np} = \frac{d_{np}}{d_{np}} \]

where \( d_{np} \) denotes the length of the lines in the trapezoid-like is discarded. This step to an accider lig of a trapezoid-like parallelism, proximi threshold values are these values can be supporting hypothe:

### 4.8 Generic Aircraft

Once the symbolic & generic aircraft modulures are derived in a the highest-level pro: This amounts to ver representing the diff highest-level product.
Qualitative Recognition of Aircraft in Perspective Aerial Images

On the other hand, a group of four lines must satisfy the following conditions:
- two lines are non-parallel, two are parallel,
- the parallels form the opposite sides of the trapezoid and so also the non-parallels,
- all pairwise line-intersections occur near detected corners, and
- the parallel lines are smaller than the non-parallel.

To overcome the problem due to oversegmentation, i.e., fragmentation of long lines into smaller parts, groups of lines that are found to satisfy any one of the above sets of conditions are merged based on collinearity measures. Non-overlapping line pairs that 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 is proportional to the smaller line length. Since, the proximity measure defined in [31] is based on the Euclidean distance between the end points of a line pair and can therefore be similar for both parallel as well as non-parallel lines, we introduce an additional perceptual measure for determining the proximal non-parallel lines only. Given the point of intersection, \( p \), of two lines \( l_1 \) and \( l_2 \), if \( p_1 \) and \( p_2 \) denote the endpoints of these lines closest to \( p \), respectively, then the end point proximity measure, \( \text{prox}_{np} \) for non-parallel lines is

\[
\text{prox}_{np} = \frac{1}{(d_{p1p} + d_{p2p})/2.0}
\]

where \( d_{ab} \) denotes the Euclidean distance between \( a \) and \( b \). Finally, if the total length of the lines in a group is smaller than a certain fraction of the perimeter of the trapezoid-like shape obtained by connecting these lines, then the group is discarded. This step is motivated by the fact that no such group of lines can be due to an accidental alignment of the constituent lines satisfying the above conditions of a trapezoid-like shape. Finally, associated with each perceptual measure of parallelism, proximity, and collinearity is a range of threshold values. Initially, the threshold values are set to the maxima of the corresponding ranges. However, these values can be relaxed based on the flow of evidence when multiple mutually supporting hypotheses interact [13].

4.8. Generic Aircraft Recognition

Once the symbolic features have been derived, these need to be matched to the generic aircraft model through an evidence accumulation process. Since the features are derived in a class specific manner, the recognition amounts to satisfying the highest-level production rule associated with the generic aircraft description. This amounts to verifying the mutual connectedness of the symbolic features representing the different parts of a generic aircraft in a manner specified by this highest-level production rule.
Instead of associating evidence with primitive features and casting these evidences in a conflict resolution situation to support or refute hypotheses individually [32] or within a "blackboard" paradigm, our effort has been directed towards gradual accumulation of evidence through

- clutter rejection,
- non-shadow feature identification,
- symbolic feature extraction, and
- interaction of hypotheses.

This reduces the number of alternative hypotheses that need to be considered at this stage. We avoid using any probabilistic measure of confidence in a hypothesis since probabilistic measures are highly context-dependent and vary widely from image to image. However, the ordering of the test conditions in the highest-level production rule are based on the likelihood of an individual test to support the evidence of a generic aircraft. One such test condition involves the most dominant structural parts of the generic model. As evidences of these structures, we make use of the dominant axes of the aircraft's shape that are extracted in the step described in Sect. 4.4.

Recognition confidence is said to be low if some of the important test conditions, called the critical evidences, of the production rule fail, such as non-identification of dominant features. The generic aircraft recognition process then needs to interact with the low-level feature regulator so that less salient symbolic features may now be available along with the existing ones. In other situations, not all the evidences associated with a hypothesis can be found, particularly when there are viewpoint-imposed constraints, such as missing structural parts due to self-occlusion. The generic recognition process then accesses deeper levels of the hierarchical database to look for alternate evidences, such as more "specialized" features, which can be verified from the data.

### 4.9 Qualitative Description Refinement

Since the generic aircraft recognition step is concerned with generating and testing hypotheses based on evidence and not on the "completeness" of symbolic information, 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 model instance which is composed of the axes of symmetry of the individual components. The skeleton can be directly used for mensuration purposes when performing quantitative matching. The final output are the identified symbolic parts of the generic aircraft.

#### 5 Experiment

The aircraft grammars are main subcor steps of the a region segm that run und UNIX progr as to provide level language the entire sy: three levels: t iprises subclasses of the models. The levels.

Example 1. Fire four C-130's.
4.10 Refocused Matching

The labeled symbolic parts which capture the global shape description of a generic aircraft are now used to direct the image-based search for more localized features that are available at lower levels of the database hierarchy. Availability of these features allows more precise classification of the recognized generic aircraft. The refocused matching process may utilize the symbolic/primitive features that 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 in C programming language on Sun Sparcstation running UNIX 4.2BSD. There are ten main subcomponents (program modules) to the system implementing the various steps of the algorithm described in Sect. 4. Of these, the modules implementing the region segmentation and dominant axes extraction are written as KBVision tasks that run under X windows. Currently, the remaining eight modules run as regular UNIX programs. However, these are also being converted into KBVision tasks so as to provide a homogeneous IU environment. A control program written in shell-level language monitors the interactions among the various modules and oversees the entire system. In our implementation, the hierarchical model database has three levels: the top-level consists of generic objects, the intermediate level comprises subclasses of a generic object, and the bottom level has the specific object models. The results reported in this paper are, however, based on the first two levels. The top level has one generic object – aircraft – and three classes at the intermediate-level – large, medium, and small. The non-image information is provided to the system in the form of an external file that contains the ancillary data: the camera-platform position/orientation, weather condition (sunny/cloudy/hazy), sun angle (if sunny), and camera parameters. The following values are used for the different thresholds mentioned in Sect. 4: $T_0 = 10$ pixels, $T_1 = 100$, $T_2 = 160$, $T_3 = 30$, and $T_4 = 0.9$, and $T_5 = 30^\circ$.

The experimental results of qualitative aircraft recognition are presented using aerial photographs ($4K \times 4K$) of an air-base. The examples are ordered according to increasing level of complexity.

Example 1. Figure 6a shows the first of the set of images which has several aircraft – four C-130's and one F-18. Using the multiresolution focusing approach, several

---

1 KBVision is a registered trademark of Amerinex Artificial Intelligence, Inc.
Fig. 6. An aerial view of an airfield: a original image (4K × 4K); b preliminary regions of interest (black regions); c new regions of interest (ROIs) found in the close-ups of the preliminary regions.

Fig. 7. Results of low-level processing of the bottom ROI in Fig. 6c: a original ROI image (162 × 240). Extraction of thinned edges using different thresholds for edge magnitude: b t = 225; c t = 200; d t = 150; e t = 100; f t = 50.

regions of interest are identified and analyzed by the object recognition system in succession. Here, we present the results of analyzing one ROI (162 × 240) from Fig. 6 that contains the F-18 aircraft. The ROI and the output of the multi-threshold edge detection step are shown in Fig. 7. In our implementation, we have selected five threshold values which are fixed for all images. As we observe, a significant part of one wing is missing from the edge image corresponding to the highest threshold value and too much of image clutter is present in the image correspond-

According to the salient str top-level (the next s mo of the pemit salient str)

The si often the f. for t = 225) set of lines/tot lines accounts fr any subset axes of regi.

Accord is cloudy, th of lines are next using t incidence o:

Figure 9a sh descriptio extraction o ezedoid-like fe
Qualitative Recognition of Aircraft in Perspective Aerial Images

The result of this step is presented in Fig. 8a which shows the top-level (globally most salient) structure, the aircraft in this case. Figure 8b shows the next set of salient structures in addition to what has already been detected, most of which are due to the mosaic of the tarmac. The following step is to extract the primitive features from this global structure. The results of line fitting to the salient structures are shown in Figs. 8c,d.

The significance of salient feature selection is summarized in Table 4. Very often the feature extraction process can lead to undersegmentation (e.g., 18 lines for \( t = 225 \)) or oversegmentation (e.g., 213 lines for \( t = 50 \)). On the other hand, the set of lines belonging to the most salient structures constitutes only 17% of the total lines obtained from the edge image of Fig. 7f. At the same time, this set accounts for nearly 84% of the "useful" (that may be associated with the aircraft) lines of the latter set. This is a significant gain in terms of computational efficiency for any subsequent model-matching step. Segmented regions and the dominant axes of regions are shown in Figs. 8e,f.

According to the ancillary data, the weather condition for the image of Fig. 6a is cloudy, therefore shadow-line removal step is skipped. Since, the convex groups of lines are also required for the symbolic features like wings, these are determined next using perceptual cues of proximity, collinearity, coterminal (based on the incidence of line pairs on detected corners) and the region segmentation results. Figure 9a shows the six convex sets of lines identified in this manner. The symbolic descriptions of some of the subparts, such as wings, tails, and rudder, require extraction of trapezoid-like structures from the convex groups. Extracted trapezoid-like features are shown in Fig. 9b. In contrast to wings or tails, subparts, such
Fig. 9. Results of qualitative object recognition: a six convex groups of lines identified in Fig. 7a; 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

as fuselage, require less specialized features like perceptual groups of collinear, parallel lines.

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 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 have been satisfied, can their ensemble be recognized as a generic aircraft. The identified subparts are shown in Fig. 9c. Inability to identify both wings or the fuselage is considered to be a recognition failure for the generic aircraft class. In this case, since both wings are detected and so also the fuselage, the recognition of a generic aircraft is successful.

The computational advantage gained because of the use of symbolic features is evident from the results of Tables 4 and 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. 7f for aircraft recognition. 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. 9d,e. Observe that the left tail of the aircraft has not yet been found, since it is not necessary for the recognition of a generic aircraft. Also, the nose-shape is not a characteristic of an F-18. The labeled symbolic features are only coarse descriptions of the object. Note that no precise model has been utilized in this recognition process.

The labeled symbolic parts are utilized by the refocused matching process which tries to perform an improved classification of the generic aircraft based on

<table>
<thead>
<tr>
<th>Table 4. Significance of the model features</th>
<th>No. of lines detected</th>
<th>magnitude threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t = 225$</td>
<td>$t = 200$</td>
</tr>
<tr>
<td>$(L_x)$</td>
<td>$(L_y)$</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. Significance</th>
<th>No. of lines from most salient structures</th>
<th>No. of distinct convex structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>
the engine location. The symbolic description of an engine is an elongated blob-like region, where the elongation axis is nearly perpendicular to the leading-edge (edge closer to the nose) of the wing when engines are located on wings, and the axis is nearly parallel to fuselage when engines are located on fuselage. The refocused matching stage interacts with the primitive feature extraction process to obtain segmented regions within a window along the leading-edges of the two wings and also the rear-part of the fuselage, separately. However, no elongated blob-like region is detected that may indicate presence of engines. Therefore, the generic aircraft is identified as belonging to a small class (Fig. 9f).

Example 2. A second aerial image is shown in Fig. 10a of which Fig. 10b constitutes an ROI containing a Hercules aircraft. The region segmentation results of Fig. 10c follow the observation that the gray-levels of the aircraft and its cast shadow are nearly similar in Fig 10b. The extracted dominant axes are also shown in Fig. 10c. In this case, three dominant axes have been determined one of which is actually due to the shadow cast by the wings. We will refer to this axis as the shadow axis in subsequent discussions.

Figure 11a shows the top-level structure obtained using the ROI of Fig. 10b. The line-fits to this structure and the next incremental salient structure are displayed in Figs. 11b,c. Once again, the computational advantage gained by using the salient structure detection step is indicated in Table 6. Here, the salient structures constitute only 26% of the lines obtained from the edge image corresponding to

| Table 4. Significance of saliency-based feature regulation for the ROI of Fig. 7a. The lines referred to in this table are those that are at least 10 pixels long |
|---|---|---|---|---|---|
| No. of lines detected using edge magnitude threshold of |
| t = 225 |
| (L₁) |
| t = 200 |
| (L₂) |
| t = 150 |
| (L₃) |
| t = 100 |
| (L₄) |
| t = 50 |
| (L₅) |
| No. of lines from most salient |
| Ratio |
| l₁/L₅ |
| Fraction of aircraft contour (ground truth) |
| 18 |
| 28 |
| 40 |
| 91 |
| 213 |
| 36 |
| 0.17 |
| 0.84 |

| Table 5. Significance of symbolic feature extraction for the ROI of Fig. 7a. The lines referred to in this table are those that are at least 10 pixels long |
|---|---|---|---|---|
| No. of lines from most salient structures |
| (l₁) |
| No. of distinct lines forming convex groups |
| (l₂) |
| No. of distinct lines forming trapezoids |
| (l₃) |
| No. of additional lines needed for recognition |
| (l₄) |
| Ratio |
| l₂/l₁ |
| Ratio |
| (l₁ + l₄)/l₁ |
| 36 |
| 18 |
| 13 |
| 5 |
| 0.5 |
| 0.5 |
Fig. 10. A second aerial image: a original image (4K × 4K); b a ROI image (300 × 450) of the aircraft marked with a × in a; c extracted dominant axes

lowest edge threshold. Table 6 also illustrates the difficulty posed by the current scenario as nearly equal number of lines 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. 11b. The potential shadow lines identified among the lines of Fig. 11b are shown in Fig. 11d. In order to resolve the shadow lines more accurately, it is first determined whether one of the axes is due to a cast shadow. This procedure is similar to matching a potential shadow line with a non-shadow line which is outlined in Sect. 4.6. In this case, the rightmost of the two nearly parallel axes of Fig. 10c is determined to be due to shadow since it matched up with the leftmost of the two when the illumination projection point information is used. Assuming that the shadow is cast on a plane, the shadow axis demarcates an area in the image away from the illumination projection point (to the right of the shadow axis) and all the potential shadow lines that lie within this area are marked as true shadow lines. For the potential shadow lines that are outside this area, the usual correspondence-based matching discussed in Sect. 4.6 is carried out. The final shadow lines are displayed in Fig. 11e. The groups of non-shadow lines are next used to obtain the trapezoid-like symbolic features of Fig. 11f.

One of the symbolic features of Fig. 11f that aligned with one of the dominant axes, axis-1 (the wing axis), is hypothesized as the wing, wing-1 (say). This is shown in Fig. 12a. (Notice that the shadow axis is removed from consideration during the generic recognition process.) Now, this generic recognition process is confronted with a situation in which only one wing hypothesis could be formed. As a result, the system accesses the lower-level of the database hierarchy for
Qualitative Recognition of Aircraft in Perspective Aerial Images

"specialized" features, like engines, that are characteristic of some aircraft, viz., large-class aircraft. Further evidence of wing-1 is sought by searching for elongated blob-like regions (qualitative description of an engine) in the vicinity of wing-1. The search succeeds in this case, thereby improving the combined support of wing-1 hypothesis. Next, the system attempts to seek evidence of the other wing, wing-2 (say), by interpreting the current situation as one of missing data.

Fig. 11. Results of feature extraction: a detection of the most salient structure; b fitting straight lines to the structure of a; c straight line-fits to the next incremental salient structure; d potential shadow lines; e resolved shadow lines; f trapezoid-like shapes identified using non-shadow groups.
Fig. 12. Results of qualitative object recognition: a a hypothesized wing, a search region for the second wing and the non-shadow lines to support it; b the current non-shadow lines together with the lines of Fig. 11c obtained from the regulator; c emergence of additional non-shadow lines within the search region of a; d recognition success after lines within the search region have been grouped by relaxing the perceptual constraints thereby allowing the other wing to emerge; e refined structural parts; f class recognition.

A search region is next set up, for this purpose, on the other side of the intersecting dominant axis, axis-2 (the fuselage axis), and along axis-1. The search region delineated in Fig. 12a is determined from the image location of wing-1 and the condition of symmetry of the wings about the fuselage axis. Next, a search for the lines that may support the evidence of the other wing is carried out within this region. However, the non-shadow lines contained within this region (Fig. 12a) fail
to identify any symbolic feature that may support a wing hypothesis. The removed shadow lines are next considered, which also are inadequate in this case. Integration of subsequent symbolic feature extraction and recognition is therefore initiated by allowing the generic recognition process to interact with the feature regulator. The new primitive features together with the current non-shadow lines that are not utilized by the symbolic features of Fig. 11f are shown in Fig. 12b. The additional non-shadow lines together with the previous ones that are identified within the search region (Fig. 12c) are further subjected to the grouping process for symbolic feature extraction. The initial (maximum) values of the various perceptual constraints fail to produce any meaningful perceptual grouping as the line primitives are few in number and are separated apart. However, since the hypotheses wing-1 and wing-2 are mutually reinforcing, i.e., evidence for one makes the evidence for the other more likely, the constraints are relaxed in steps. Particularly, the lowering of the thresholds of the proximity of a line-pair intersection to a detected corner and the ratio of the total line length to the perimeter of a trapezoid-like shape (see Sect. 4.7) 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.

The identified structural parts are the two wing tips, the front part (and a small portion of the rear) of the fuselage, and the nose. These are indicated in Fig. 12d and the refined parts are shown in Fig. 12e. The processes of shadow identification and symbolic feature extraction have reduced the number of useful lines for aircraft recognition to only 6% of the total number of lines (Table 6) that otherwise would have to be considered. Finally, the refocused matching process determines the class of this aircraft as large as indicated in Fig. 12f.

<table>
<thead>
<tr>
<th>No. of lines from the most salient structures ( (l_i) )</th>
<th>Fraction of ( l_i ) from ( E_{90} ) in set ( l_i )</th>
<th>Fraction of ( l_i ) belonging to aircraft (ground truth)</th>
<th>Fraction of ( l_i ) belonging to shadows (ground truth)</th>
<th>No. of shadow lines found ( (l_j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>0.26</td>
<td>0.38</td>
<td>0.35</td>
<td>36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of additional lines from the next salient structures ( (l_j) )</th>
<th>No. of distinct lines forming trapezoids ( (l_i) )</th>
<th>No. of additional lines needed for generic recognition ( (l_j) )</th>
<th>Ratio ( (l_i + l_j)/(l_i + l_j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>6</td>
<td>5</td>
<td>0.06</td>
</tr>
</tbody>
</table>
6 Conclusions

In this chapter, we have presented an approach to qualitative recognition of aircraft in complex, perspective aerial imagery. 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. We have introduced a framework for geometric model-based object recognition that addresses some of these variabilities of the real-world and also the fundamental issues related to model-based recognition. In particular, our approach is capable of handling shadows, clutter, and low image contrast and it uses qualitative features of object (aircraft) models for recognition. 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 extraction ones. Real-world data, highlighting the difficulties of aircraft recognition in practical situations, is used to demonstrate the effectiveness of our proposed approach. Implementing the final stage, that of quantitative recognition, is a future research issue.

A Generalized Cylinders for Model Representation

In this section, we discuss the properties of Straight Homogeneous Generalized Cylinders (SHGCs) that are used to model the volumetric shape of generic aircraft and also the shadow geometry. Subsequently, we present analysis of the occluding contours of these shape descriptors, the results of which are used directly in our algorithm.

A.1 Contour Analysis for SHGCs

In this subsection, we shall describe the properties of the occluding contours, i.e., the contours tangent to the line of sight, of SHGCs using some of the results of Shafter [45]. The occluding contour is a curve on the surface of an object that projects to the silhouette of the object in the image.

Recall that an SHGC is a subclass of Generalized Cylinder (GC), a shape descriptor of a large class of curved objects. It has a line segment in space for its axis and an uniformly-scaled cross-section along this axis. Figure 13a shows an SHGC which maps two parameters onto a set of points in the 3-D world: $s$, the parameter which measures distance along the axis of the shape; and $t$, the para-
Fig. 13. Straight Homogeneous Generalized Cylinder: a representation, and b coordinate axes.
Figure 14. Straight Homogeneous Generalized Cylinder: a coordinates of a point on the surface, and b surface normal at a point.

Fig. 15. Perspective projection of an SHGC onto an image plane where O is the viewpoint. \( \rho = (\rho_x, \rho_y, \rho_z) \) is the view vector in the \( w - v - s \) coordinate system.

\[
W(t) = \begin{vmatrix}
\frac{du}{dt} & \frac{dv}{dt} \\
\frac{dv}{dt} & \frac{dw}{dt}
\end{vmatrix}
\]

Figure 15 shows the perspective projection of an SHGC onto an image plane, in which the contour of the silhouette is the projection of the occluding contour. Since along the occluding contour the surface is tangent (i.e., the surface normal is perpendicular) to the viewing direction, the following relation is satisfied.

\[
N \cdot \rho = 0.
\]
Qualitative Recognition of Aircraft in Perspective Aerial Images

Now, in the case of a Right SHGC (RSHGC), $\alpha = \pi/2$, and for a Linear Homogeneous Generalized Cylinder (LHGC), $r = m(s - s_o)$, where $s_o$ is the apex and $m$ is the gradient of the linear envelope. Thus, for a linear RSHGC, the occluding contour satisfies

$$\rho_o \frac{dv_c}{dt} - \rho_s \frac{du_c}{dt} - m\rho_o W(t) = 0. \tag{9}$$

Given an object whose shape is represented by a linear RSHGC, the viewing vector $\rho$ (i.e., $\rho_o, \rho_s, \rho_p$), and the parameter $m$ are fixed. Hence, (9) is a function of $t$ only.

**Lemma 1.** The occluding contour of a Linear RSHGC always lies in a plane.

Proof. Let (9) hold true for some value of $t = t_o$ and let $P(s, t_o)$ be any arbitrary point on the occluding contour such that

$$P(s, t_o) = (w_p, v_p, s_p) = (u_c(t_o)r(s), v_c(t_o)r(s), s).$$

The corresponding normal is

$$N(s, t_o) = r(s) \left( \frac{dv_c}{dt}, -\frac{du_c}{dt} - mW(t_o) \right).$$

We observe that

$$r(s) \left[ (\rho_o - w_p) \frac{dv_c}{dt} - (\rho_s - v_p) \frac{du_c}{dt} - mW(t_o)(\rho_s - s_p) \right]$$

$$= r(s) \left[ -w_p \frac{dv_c}{dt} + v_p \frac{du_c}{dt} - ms_p W(t_o) \right]$$

$$= r(s) \left[ -u_c(t_o)r(s) \frac{dv_c}{dt} + v_c(t_o)r(s) \frac{du_c}{dt} + msW(t_o) \right]$$

$$= r(s) \left[-r(s)W(t_o) + msW(t_o)\right] = 0 \tag{10}$$

for $s_o = 0$. But (10) is the equation of a plane through the occluding contour and containing the viewpoint.

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

Proof. The projection of a plane when viewed end-on is a line. The tangent plane passing through the viewpoint and containing the occluding contour of a Linear RSHGC therefore projects to a line which is the silhouette corresponding to the occluding contour.
Lemma 3. The 2D image of a solid corner subtended at the apex of a Linear RSHGC or formed by two intersecting Linear RSHGCs is a corner.

Proof. A solid corner at the apex of a Linear RSHGC or at the intersection of two Linear RSHGC is due to the dihedral angle between two intersecting tangent planes of the occluding contours. Now, each of the tangent planes projects as a line in the 2D image according to the above claim, and two lines corresponding to the two intersecting tangent planes subtend a corner between them in the image.

A.2 Shadow Models

In this subsection, we discuss the properties of shadow formation geometry which is modeled using SHGC shape descriptors. The shadow of an object has two parts: the object's dark side (or self-shadow) and its projected (or cast) shadow. The shadow boundary is defined by a brightness discontinuity between directly illuminated regions and regions receiving either no light or only indirectly scattered light. The apparent boundary is completely determined by the surface topography and the positions of the light source and the viewer.

A shadow boundary has three basic types of segments [24]: shadow-making, shadow, and occluding. These segments for an object illuminated by a point source are illustrated in Fig. 16. Each segment represents a certain profile of the object. The shadow-making boundary is essentially the occluding contour for the point source. For an object whose shape is represented by a Linear RSHGC, the shadow-making boundary is planar. The “shadow volume” (the volume of space shaded by an object) is an LSHGC with the point source as the apex. This may be seen from Fig. 16. The end of this LSHGC closer to the source is made up of the plane containing the shadow-making boundary.

Lemma 4. The shadow boundary on a plane for a Linear RSHGC is linear.

Proof. We note that the shadow boundary is the intersection of the tangent plane containing the shadow-making boundary and the plane on which the shadow is cast.
References


Image Technology
Advances in Image Processing, Multimedia and Machine Vision

With 344 Figures, Some in Color and 41 Tables

Springer