[26] The Design of the IMAGINE II Scene Analysis Program

Robert B Fisher

Department of Artificial Intelligence
University of Edinburgh, Edinburgh EH1 2QL, UK

Reprinted, with permission of John Wiley and Sons Ltd, from *From Surfaces to Objects: Computer Vision and Three Dimensional Scene Analysis* 1989.

1. Introduction

The IMAGINE I [1] system was designed for scene analysis in a laboratory or factory domain. The scenes will contain multiple overlapping man-made, but non-polyhedral, objects. Its inputs were segmented surface patches with associated range and surface orientation measurements, and surface-based hierarchically structured object models. Experience with this program has led us to the complete redesign embodied in the IMAGINE II system.

While the re-implementation is complete for only a few of the modules now, the design of the system is complete. This brief paper summarizes the key features of the design, to help place some of the other papers in this book and their motivations into context.

1.1 Critical Review of IMAGINE I

IMAGINE I [1] was designed for scene analysis starting from a labeled, segmented 2 1/2D sketch. That is, it used a surface-based description of the scene, where the surfaces were segmented into regions of nearly constant curvature class, and the boundaries between the regions were labeled with the discontinuity type (i.e. whether a depth, orientation or curvature discontinuity).

From this, individual surface patches were joined or extended across occlusion boundaries, to deduce as much of the original surface shape as possible.

Next, surface patches were grouped to form surface-clusters [2], which were an indentity-independent volumetric scene representation. The purpose of the representation, in the context of 3D scene analysis, was to create contexts within which to accumulate or find evidence for model invocation and hypothesis completion. That is, when working with a given model, only evidence from within the context would be used.

Three-dimensional properties of the surfaces and surface clusters were then estimated (e.g. surface curvatures and areas) using the 3D information present in the data.

Model invocation [5] then occurred. To allow:

- complete access to the whole model base,
- efficient, non-directed computation over all models,
- integration of structural and generic as well as property evidence, and
- graceful degradation as property values become unreliable or missing

the model invocation process was formulated as an evidence accumulation computation evaluated in a network suitable for wide-scale parallelism.

The object models were based on quadratic surface patches linked together to form laminar or solid assemblies, which could then be used to hierarchically construct larger assemblies. Variables could be included in the reference frame transformations.

Model matching was initiated by the invocation process and occurred between models and data features found in the surface or surface cluster contexts. The matching process used the geometrical models to deduce a reference frame for the object, and from this deduced which features of the object were likely to be invisible, partially obscured or fully visible. Then, it searched the image data for evidence justifying all the unlocated visible features (or evidence for their absence, such as being obscured by other objects).

The key example scene analysed was that of a PUMA robot with its gripper obscured by a trashcan. The program successfully deduced the identity of all subcomponents, was largely correct in its visibility analysis and made reasonable estimates of all components (including deducing the joint angles between the links of the robot).

Some of the weaknesses of IMAGINE I are identified here, and provide the impetus for the IMAGINE II design:

- Only reasonably complete surface patches could be used (i.e. no patch fragmentation).
- Data could have some numerical errors, but completely erroneous data (e.g. boundary mis-labeling) would thwart successful scene analysis.
- Object models used only surface patches and these could not extend around the model. Holes and

non-surface evidence were not modeled.

- Data contexts were based only on surface information and did not exploit or account for curve and volumetric data.
- Model invocation did not correctly account for generic relationships and did not exploit spatial configuration evidence.
- Model matching was primarily bottom-up and required complete surface patches.
- The geometric reasoning was based largely on transforming and intersecting parameter ranges represented as 6D parameter rectangular solids, resulting in weak parameter estimation.

2. Overview of the IMAGINE II Design

Figure 1 shows a block diagram of the main modules of the IMAGINE II system. As in IMAGINE I, this program assumes the data comes from segmented 2 1/2D sketch-like data, only here curve and volumetric scene features may be part of the input. Also, the data is allowed to be fragmented and possibly incomplete. The input data structure is a REV graph (Region, Edge, Vertex). In the context of Alvey consortium II, the REV is instantiated using data from the Sheffield GDB system (elsewhere in this book). Alternatively, the data might come from segmented laser ranging data.

The system output is, as before, a list of object hypotheses with position and parameter estimates and a set of image evidence and justifications supporting the object hypothesis.

The rest of this section summarizes the design and some of the ideas behind the modules and data structures shown in figure 1.

2.1 Building the VSCP Structure

The first new representation is the VSCP structure (Volume, Surface, Curve, Point), which is constructed from the REV by knowledge-based structure completion processes. The goal of this process is to group curve and surface features from the REV to overcome fragmentation and occlusion effects and to remove non-structural artifacts (e.g. reflectance edges). It is possible that the original raw data might be interrogated to help verify deductions, but this is not planned for at present.

An example of an occlusion rule is:

- If two valid "TEE" junctions lying on the boundary of the same surface can be extended (using the local boundary shape) until they intersect, and the curve extensions lie behind closer surfaces, then hypothesize that the original shape of the partially obscured surface is that of the extended surface.

An example of a fragmentation rule is:

- If two surface patches are "adjacent", have similar shape, depth and orientation and there are not intervening space curves (e.g. from patch edges or closer surfaces), then merge the two patches.

Here "adjacent" is a heuristic concept because the surface characterization is assumed to be neither complete nor dense (i.e. there may be missing surfaces and there might be substantial gaps between adjacent patches).

2.2 Building the Contexts Structure

Invocation and matching will still occur in data contexts, only now contexts are provided for curve and volume hypotheses as well as surface and surface clusters. The point of these structures is three-fold:

(1) They improve matching efficiency by grouping related data and thereby isolating irrelevant data.

(2) They create a structure that can accumulate plausibility for model invocation.

(3) They represent the 3D scene structure in an identity-independent manner and support vision-related processes such as autonomous vehicle navigation or robot grasping.

Points (1) and (2) are most relevant here.

The context structures are hierarchical in that contexts can be grouped to form larger contexts. Contexts are designed to support recognition of curves, surfaces, volumes and larger assemblies of features, so one context type exists for each. For example, the information contained in an surface context might link to both curve fragments and surface patches, because either might help define a complete surface.

Examples of context-forming rules are:

- If a set of adjacent surface patches are completely isolated by depth discontinuity boundaries and there are no such boundaries internal to the group, then these surfaces form a context for recognizing an assembly.
- If a set of space curves roughly surrounds a region of 2D image space and the curves are not radically different in depth, then hypothesize a surface context lies within the curves.

2.3 Structure Description

Model invocation and hypothesis completion require property estimates for image features. Because we are using 2 1/2D sketch data, 3D properties can be directly measured. These properties may be associated with isolates features, such as:

curve fragment properties: length, curvature, ...

surface fragment properties: area, curvature, elongation, ...

or they may be associated with pairs or groups of features, such as:

curve fragment pairs: relative orientation, relative size, ...

surface fragment pairs: relative orientation, relative size, ...

While this project mainly considers structural properties, this module could also attach non-structural properties, such as colour, gloss, texture or surface markings.

2.4 Model Invocation

Model invocation occurs roughly as in IMAGINE I (section 1.1). A network implements the computation in

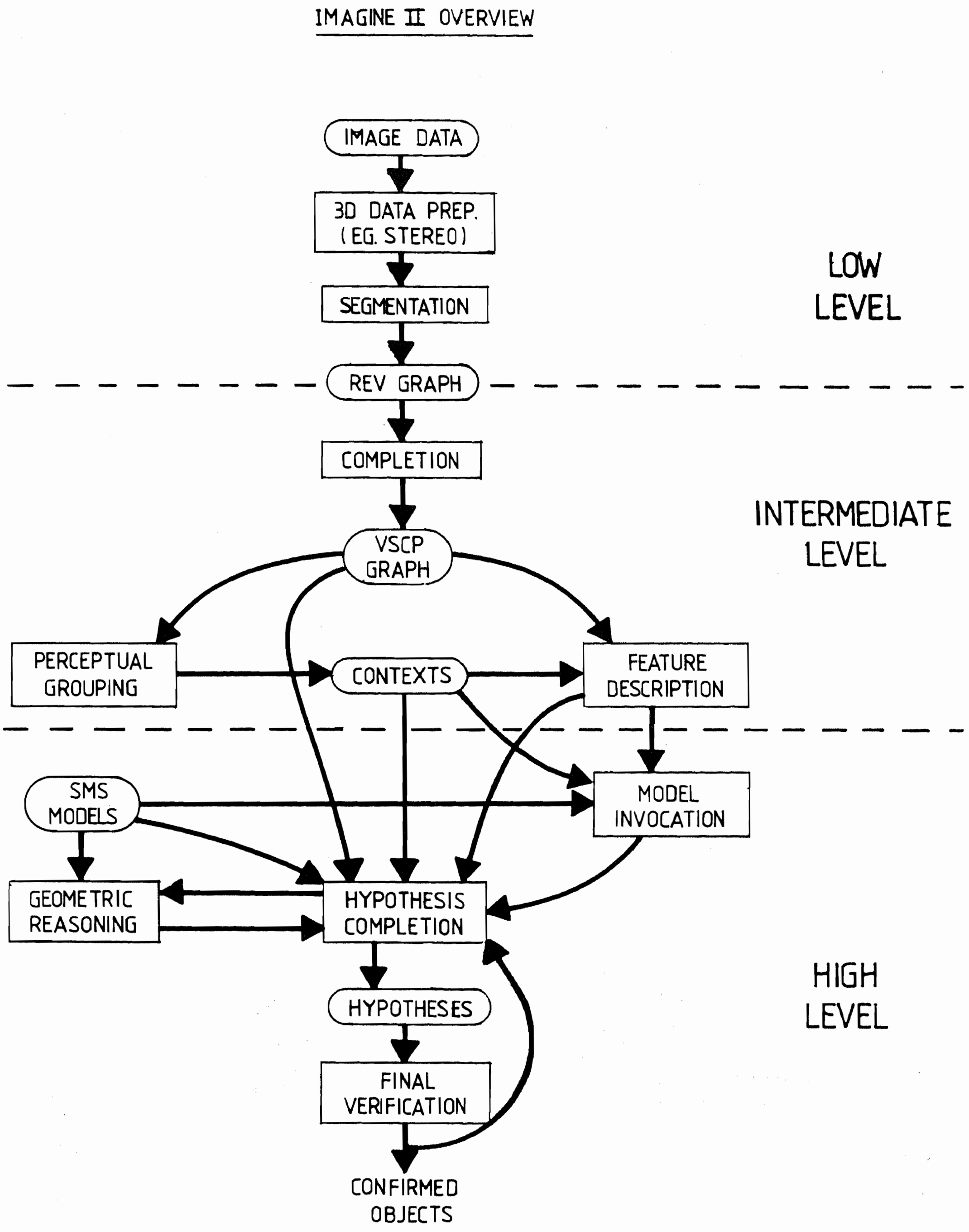

Figure 1 - Structure of the IMAGINE II System

a manner suitable for parallel evaluation. Nodes represent the pairing between individual model and data features, and are connected to other nodes according to the type of relation. Relations include: structural (e.g. "subcomponent of"), generic (e.g. "visual specialization of"), class (e.g. "non-visual specialization of"), inhibiting and general association. Invocation occurs as a result of a plausibility computation, where plausibilities arise from direct evidence (e.g. from a measure of the fit between data and model properties) and indirect evidence imported from the related nodes.

Improvements over IMAGINE I include those of Paechter [8] relating to generics, symbolic properties, property evaluation and uniform integration of direct evidence types. Other extensions include: property evidence from spatial configurations and data feature re-use inhibitions.

2.5 Object Models

The object models used are the SMS models [4,6], as described in a companion paper in this book ("SMS: A Suggestive Modeling System for Object Recognition").

The SMS models are primarily structural with model primitives designed to match with either curve, surface or volumetric data as alternatives. The models are hierarchical, building larger models from previously defined substructures. Substructure placement is by local reference frame transformation.

All model dimensions and reference frame transformations may involve variables and expressions, and algebraic constraints can bound the range of the variables.

A generic hierarchy can be constructed to embody both scale-based and abstraction simplifications.

An important part of the models are the viewpoint-dependent feature groups, which record the fundamentally distinct viewpoints of the object. They also identify model features visible from the viewpoint and identify new viewpoint dependent features (such as occlusion relationships or extremal boundaries).

2.6 Hypothesis Completion

Initial selection of the model may come bottom-up from invocation or top-down as part of another hypothesis being completed. Hypothesis completion then attempts to find evidence concerning all model features.

Feature visibility information comes from selection of a viewpoint-dependent feature group from the SMS model (section 2.5) as selected according to the estimated orientation of the model (from geometric reasoning - section 2.7). This will inform on the visibility of most model features (i.e. whether tangential, backfacing, self-obscured, etc.).

Completion is largely a hierarchical synthesis process that groups recognized subcomponents to form larger hypotheses. The most primitive features are designed to be recognized using either curve, surface or volumetric data, depending on what is available. At all stages, geometric consistency is required, which also results in more precise position estimates and estimates for embedded variables (such as a variable rotation angle about an axis).

Completion is a heuristic process whereby various approaches are tried to find evidence for a feature. For example, some heuristics for surface finding are:

(1) Use a reasonable image patch if it is in the predicted position, with the predicted orientation and has the correct shape and size.

(2) Use a smaller image patch if it is in the predicted position, with the predicted orientation and has the correct shape and no patch of the correct size is found (i.e. expect fragmented patches).

(3) Ignore the surface if it is small and far away.

Application of the heuristics is controlled through routines that know what approaches are available for finding features (and when to try them). Process management uses a task agenda (section 2.8).

2.7 Geometric Reasoning

The geometric reasoning [3,7] is described in detail in a companion paper in this book ("Geometric Reasoning for Computer Vision").

The geometric relationships between model features, model and data pairings and *a priori* scene knowledge are fundamentally represented using algebraic equalities and inequalities.

Algebraic expressions are expensive and difficult to manipulate, so to acquire estimates of object positions, for example, the set of inequalities is transformed into networks expressing the computational relationships between the variables contained in the constraints. These networks have the side-benefits of:

- improving on the first-order bounding methods through iteration and
- having a naturally parallel structure

Analysis of the types of geometric relationships occurring in scene analysis showed that most relationships could be expressed using only a small (5-6) set of standard relationships (e.g. a transformed model point maps to a data point). The standard relationships could then be used to construct standard network modules, which can then be allocated and connected as needed when solving larger problems.

By analyzing the possible pairings of SMS model features to 2 1/2D sketch features, a catalogue of network modules has been developed, to be compiled once, then allocated and connected appropriately for any given model-to-data pairing. This promotes convenient geometric testing and parameter estimation during hypothesis completion.

2.8 Agenda Management

To facilitate experimentation with different control regimes, the hypothesis completion processes are activated from a priority-ordered agenda. An agenda item embodies a request for applying a specified hypothesis completion process on a given datum or hypothesis. The activated process may then enter other requests into the agenda. We use the agenda to implement a mixed control regime involving both top-down and bottom-up hypothesis completion.

2.9 Hypothesis Verification

Because data can be fragmented or erroneous, object hypotheses may be incomplete. Further, spurious hypotheses may be created from coincidental alignments between scene features. Hypothesis completion with geometric reasoning will eliminate some spurious hypotheses, but some instances of global inconsistency may remain, such as when three unconnected planes at right angles may match a cube.

This module is not designed yet, but is intended to consider 2 problems:

(1) global consistency of evidence (e.g. connectedness and proper depth ordering of all components)

(2) heuristic criteria for when to accept incomplete models.

3 Conclusions

The system design given here is intended to cope with fragmented and somewhat incorrect data deriving from scenes containing self and externally obscured complex, non-polyhedral manmade objects including possible degrees of freedom (e.g. robot joints). The test scenes we are planning to use involve both stereo data (from the Alvey "widget" and PUMA robot scenes) and laser ranging scenes (the "oilcan", "lightbulb", "renault part" scenes, ...). While some components of this design are still in the exploratory stage, others are sufficiently developed that parallel implementations can be investigated, leading to eventual re-implementation for real-time efficiency.

This book is being written before all modules have been implemented, integrated and tested; however, it is hoped that the character of the planned IMAGINE II system is clear.

References

[1] Fisher, R. B., "From Surfaces to Objects: Recognizing Objects Using Surface Information and Object Models", PhD Thesis, University of Edinburgh, 1986.

[2] Fisher, R. B., "Identity Independent Object Segmentation in 2 1/2D Sketch Data", Proc. 1986 European Conference on Artificial Intelligence, pp148-153, July 1986.

[3] Orr, M. J. L., Fisher, R. B. "Geometric Reasoning for Computer Vision", Image and Vision Computing, Vol 5, No 3, pp233-238, August 1987.

[4] Fisher, R. B., "SMS: A Suggestive Modeling System for Object Recognition", Image and Vision Computing, Vol 5, No 2, May 1987. Also University of Edinburgh, Dept. of A.I. Working Paper 185, 1985.

[5] Fisher, R. B., "Model Invocation for Three Dimensional Scene Understanding", Proc. 10th Int. Joint Conf. on Artificial Intelligence, pp805-807, 1987.

[6] Fisher, R. B., "Modeling Second-Order Volumetric Features", Proc. 3rd Alvey Vision Conference, pp79-86, Cambridge, 1987.

[7] Fisher, R. B., Orr, M. J. L., "Solving Geometric Constraints in a Parallel Network", Proc. 3rd Alvey Vision Conference, pp87-95, Cambridge, 1987.

[8] Paechter, B., "A New Look At Model Invocation With Special Regard To Supertype Hierarchies", MSc Dissertation, University of Edinburgh, 1987.

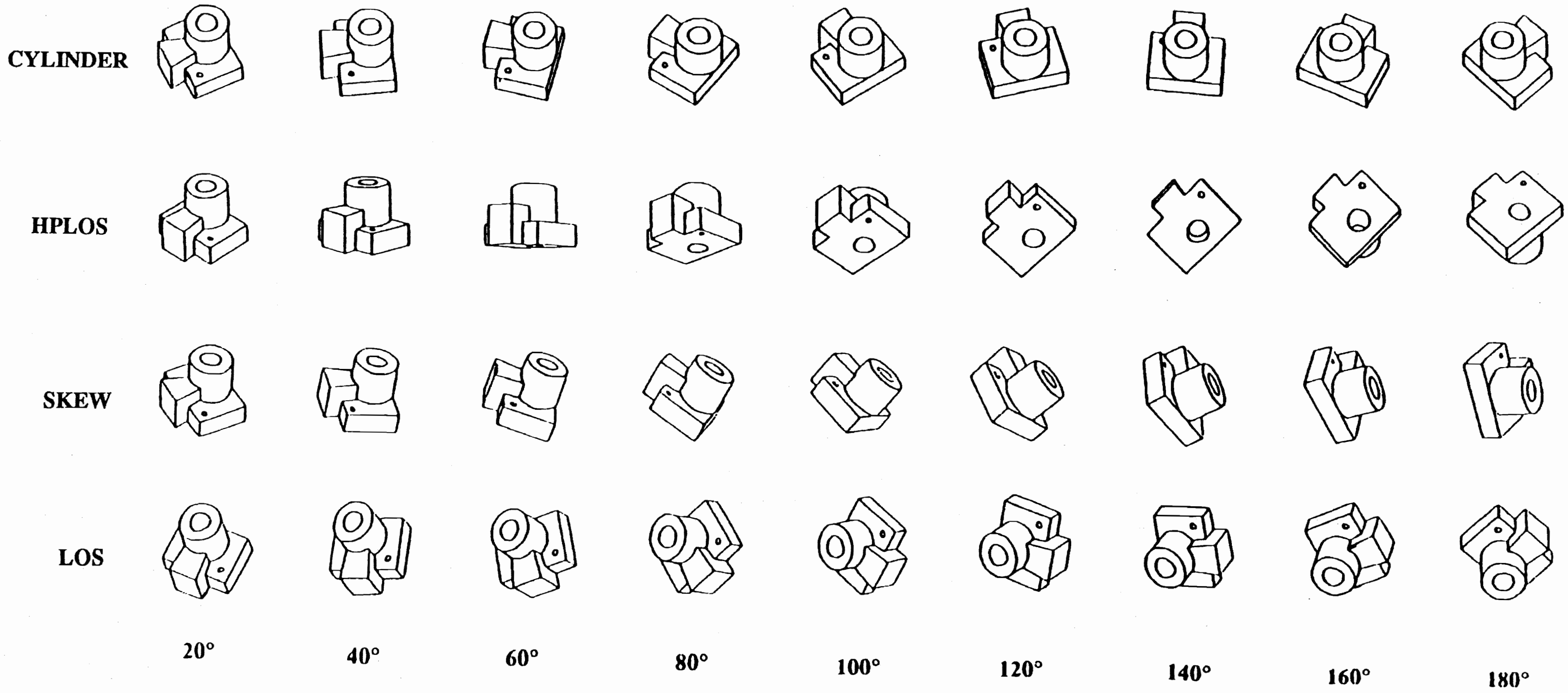

Figure 1. The complete set of rotated views used in the experimrnt arranged by axis and increment of rotation around the axis. Each rotated view was paired with the reference view (bottom).