DEPARTMENT OF ARTIFICIAL INTELLIGENCE
UNIVERSITY OF EDINBURGH

DAI Working Paper No. 144
Date: May 1982
Revised: July 1983

Title:

Reasoning About Anomalous Data in the Image Analysis Process - a Proposal

Author:

Robert B. Fisher

Abstract:

One current problem preventing effective image analysis is the pervasiveness of anomalies, such as obscuration or low contrast, and the effects they have on the analysis process. Associated work has shown that the anomalies are typical aspects of images, appearing at specific places in the information path of the image formation and analysis process, and generally producing specific effects in the results of processing. This paper suggests how, given meta-knowledge of the analysis process and the anomalies, extracted information and object models, a program can reason about apparent instances of anomalies, to both explain and correct them or suggest alternative reasoning methods. This paper is a PhD thesis project proposal, and includes discussion of related research, possible problems and the general plans for the project. Included is a revision section that discusses the specific proposal for handling obscuration.

Acknowledgements:

The author was appreciatively supported during this work by a postgraduate studentship from the University Of Edinburgh. Special thanks also go to R. J. Beattie, W. Caplinger, J. Hallam, J.A.M. Howe, and B. L. Yin for their discussion and advice on both the content and exposition of this paper.

1. Introduction

Most image analysis programs require substantial hand tuning of external (e.g. lighting) and processing (e.g. threshold) parameters before they are capable of analyzing any particular image. Then, if some aspect of the scene is changed, the program often fails to work. In retrospect, it is often clear why the particular instance failed, and the blame is placed on an anomaly (e.g. obscuration or low contrast across a particular boundary). Unfortunately, anomalies are ubiquitous.

A previous paper ([Fis83]) discussed some analysis of the anomalies that occur in the image formation and analysis process. It suggested that they were typical aspects of typical images, and that they needed to be accounted for, as well as explained, as part of the analysis process. The anomalies were discussed in sufficient detail to conclude that their effects appeared at specific locations in the information flow (of the image formation and analysis process) and that specific anomalies typically produce specific effects.

As a result of the examination, it was concluded that it could be possible to construct an intelligent analysis program that used this knowledge. In particular, this knowledge was a form of meta-knowledge, of both the normal and abnormal outputs (anomalous data) of the process. If one could identify instances of anomalous data, meta-knowledge could hypothesize the possible sources for that data. These hypotheses could then be tested, with the results used to correct erroneous data or add missing data.

This paper is a (PhD thesis) proposal for a project to study these ideas in greater detail, and then apply them to image analysis. The method is to be based on a theoretical study of:

1. the types of anomalies encountered in the image formation and analysis process,
2. methods for determining when anomalous data is present
3. methods for determining the nature of the anomaly, when encountered, and
4. methods for overcoming the anomalies.

The results of this investigation will be formalized in a computer program, to determine whether the resulting system is capable of more powerful and reliable recognition. The novelty of this approach is partly the collected meta-level knowledge of the anomalies and their effects, and partly the explicit recognition, declaration and overcoming of the anomalies, which is an integral part of interpreting a scene. Since recognition is always based, in varying extents, on hypotheseses supported by indirect evidence, the program will also record explicit justifications for its reasoning.

The research proposed in this paper will be implemented in the context of the author's **IMAGINE** image matching program ([Fis82]). This program executes under the assumptions:

all reasonable matches should be suggested and pursued,
the underlying semantics of the structures matched will prune locally incompatible matches, and
scene and object consistency will prevent globally inconsistent matches.

It is felt that the natural prunings and match failures will make a sophisticated search algorithm unnecessary.

These assumptions are implemented by matching structures according to rules defined over object features, with attached semantic routines evaluating the details of the match. This rule based framework is ideal for implementing the anomaly based reasoning because:

- the anomaly (and normal case) reasoning methods are largely independent means of deriving the same structures given different data requirements,
- the independent rule formulation allows incremental improvements, and
- the generation of all potential matches allows parallel pursuit of the final recognition.

Section 2 of the paper gives more information about the anomalies encountered in a typical laboratory setting. Section 3 describes the intended project. Section 4 critiques research on coping with scene difficulties. Section 5 overviews other research in topics on which this research is dependent. Section 6 covers the administrative issues, such as project stages, requirements, and a potential schedule. Section 7 details the expected problems of the project, and section 8 concludes with a discussion of the project's originality. Section 9 extends the proposal to cover the specific anomaly of obscuration.

2. The Anomalies

A previous paper ([Fis83]) discussed the notion of anomalies in the image formation and analysis process, and defined an anomaly as a situation in the scene or a processing event that caused the analysis process to fail. The factors responsible for the anomalies are: scene objects, scene construction, illumination, image formation, object modelling, and processing algorithms. This was then followed by an examination of the types of anomalies, the particular factors affecting them, and their results. Some of the types included: shadows, obscuration, self-obscuration, coincidental alignments, scale, model omissions, model simplifications, surface properties, noise and algorithm resolution limits. Typical results included: missing, improper, extraneous or false segmentations, mis-classified segments, missing, incorrect, ambiguous or unidentifiable structures, duplicate recognitions or improperly matched subcomponents. Proceeding from this analysis, a schema of the image formation and analysis process showed at what points the anomalies first affect the information flow.

It was felt that this analysis amounted to the beginnings of a meta-level analysis of the image interpretation process, and that an intelligent image analysis process could take advantage of this knowledge. In particular, it could use the presence or absence of certain results, with its knowledge of the processes applied in their creation to correct or augment those results. Another possibility would be to suggest alternative, secondary reasoning approaches.

Another point that was made was that the anomalies were a normal aspect of typical images and their processing, and that a good image analysis program would have to not only actively cope with them, but also explain them as aspects of the image.

3. The General Proposal

The author believes that, given:

- the type of structure being searched for,
- the prediction of the object model, general knowledge of surfaces, light sources and imaging
- the current knowledge of the scene,
- knowledge of the reasoning or processing used, and
- meta-level knowledge of the sources and effects of anomalies

it is possible to greatly reduce the set of potential causes of an anomaly, to the point where only a few tests are sufficient to identify the specific cause. Then, a special test can be applied to override the effect of the anomaly. This may consist of determining the presence or absence of a structure not found by the more general recognition or segmentation processes (or to eliminate "extra" structures, such as highlights, shadows). If the problem is such that a detection test is not feasible, as in obscuration, then the missing structure may be hypothesized, with invalid hypotheses being eliminated in the future. For example, an obscuring object (when in the correct location) allows the hypothesis of the obscured structure (given that other structures have been found). This may be strengthened by the detection of partial edges or surfaces. This would, in effect, add secondary case reasoning to the rule base of primary methods.

It is also possible for the reasoning to suggest external solutions to the problems, such as moving the camera to a more advantageous location (depending on the problem), or moving or changing the lighting, though this would require three dimensional knowledge of the scene. In essence, it could suggest adjustments in the manner of a human observer reacting to difficulties (though differently).

More formally, the three types of reasoning can be stated as:

Type I errors (present, erroneous results):

IF some structure S is hypothesized (object instance) to be a result of an anomaly during the image formation or analysis process
AND there are reasons to believe it doesn't exist at location L (incomplete object schema instance, current scene analysis)
AND current information suggests a set of possible anomalies (a) (processing model, current scene analysis, anomaly models)
THEN apply specific tests $F_a(S,L)$ and deny the existence of S if the test succeeds.

Type II errors (missing, correct results):

IF some structure S is not hypothesized (object instance) as a result of an anomaly during the image formation or analysis process
AND there are reasons to believe it exists at location L (object model)
AND current information suggests a set of possible anomalies (a) (processing model, current scene analysis, anomaly models)
THEN apply specific tests $T_a(S,L)$ and assert the existence of S if the test succeeds.

Type III errors (present, correct, unaccounted-for results):

IF some structure S is hypothesized (data instance) to be a result of an anomaly during the image formation or analysis process
AND there are reasons to believe it does exist at location L (detail, segmentation process)
AND current information suggests a set of possible anomalies (a) (processing model, current scene analysis, anomaly models)
THEN apply specific tests $D_a(S,L)$ and make hypotheses that account for the detail if the test succeeds.

Several of the anomalies are easily detected. Self-obscuration of the backside of an object is deducible. Though the model predicts a surface, when rough location and orientation information about the object is known, then it can be reasoned that the backside surface will probably not be visible. Or, if an edge is expected and two roughly contiguous tracked edges are found, terminated by two T-junctions (see figure 1), then the edge can be hypothesized to exist, but be obscured by a closer surface. If tracking a complete edge and it fades out because of noise, then a system could do a contrast test across the edge and then use an optimizing tracker or an averaging detector, or simply hypothesize.

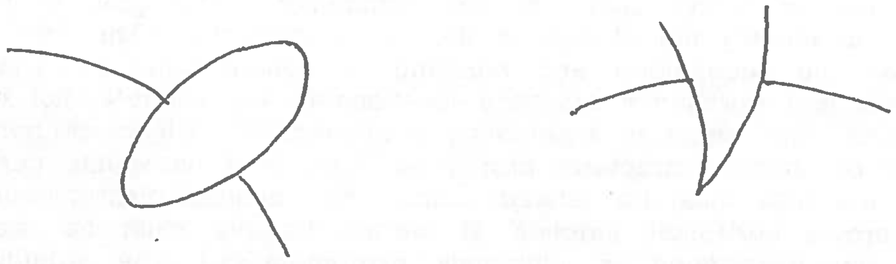

Figure 1 - edge obscured by region. Note that the "tee" junctions suggest obscuration of continuous boundaries

It is noted that the existence of the anomaly itself is often the clue to its cause (i.e., certain symptoms have clear causes). For example, a sharing of a boundary by two object hypotheses may suggest two closely aligned boundaries. Or, duplicate recognition of an object, in the same location, but with a different subcomponent, suggests alignments, insufficient discrimination or reasoning failures. In these cases, the reasoning can act on either the commonalities or differences between the structures representing the recognized objects. With two effectively indiscriminable objects, the program could suggest discriminating evidence and hypothesize why it's missing or spuriously incorrect.

It is recognized that some tests are not feasible until sufficient information has been obtained, such as scale testing without rough object sizes, or self-obscuration without object orientation estimates or shadow analysis without lighting and object orientation estimates. These problems are related to the inference of three dimensional structure from two dimensional evidence. As this is still an active research area, reasoning about these anomalies is likely to be less effective. A practical difficulty of the project is obtaining this information without incurring substantial unrelated research problems.

If there were a set of these anomaly detection and correction processes, and if they can be generalized to function over a class of models rather than just specific objects, then this would be the basis for a recognizer capable of operating in more realistic conditions.

As the rules for detecting and correcting anomalies are independent of each other, it is reasonable to embed them in a rule-based image recognizer such as **IMAGINE** ([Fis82]). This recognizer does bottom-up model-driven (procedural) exhaustive image matching. It operates under the assumption that all reasonable alternatives should be suggested and that spurious hypotheses will be pruned by failure to match at successive levels of structure. The recognizer supports independent entry and change to the rules, thus allowing progressive refinement of its operation. These rules would be implemented as specific procedures embodying knowledge of the anomaly and its potential corrections. They, however, would reason from the object and anomaly models, when doing any reasoning. Thus, they will be capable of specific reasoning about the particular type of problem, but in general over the possible entities of the domain. (This would be an effective combination of both procedural and declarative knowledge of the problem [Win75].)

To implement the anomaly based reasoning, it will be necessary to limit the range of scenes and objects recognized. For this project, the major man-made objects found in typical office or laboratory settings - tables, workbenches, chairs, robot manipulators and people will be the recognized objects. These are all objects of reasonable size, and are not all strictly rectangular or rigid. As the vision system will be model driven, the scene could contain a variety of objects, whose recognition was limited only by the modelling system and matching ability of the recognizer. The goal of the recognition will be to identify the objects in the scene correctly. The identification will include the recognition and labelling of subcomponents ("There are flashlight, robot and workbench at these locations in the scene"), but not to explain the scene ("The robot is assembling a flashlight"). Object recognition will be based on defined structural properties, (i.e., what its visible components are and how they must be related) rather than defined discrimination properties (e.g., brown horizontal patches at certain heights must be table tops). (Though any description is ultimately discriminative.) The primitive symbolic data for the recognition process will be based on some form of edge and surface patch data. The raw image will also be available for specialized tests.

The substance of the project plan can be summarized as:

Completion of the study of anomalies and their properties
Elaboration of a set of rules for the detection of, recognition of, and possible solution to a subset of the anomalies present in images. These rules will be formulated to reason about scenes independent of any particular type of scene or object model.
Generalization of these rules to cover classes of objects, rather than specific instances
Formalization of the rules as a set of programs that reason about anomalies based on specific object model and scene properties
Embedding of these programs in the context of **IMAGINE**
Demonstration of their effectiveness in both specifically tailored situations and general coincidental scenes

4. Directly Related Research

There are several areas of previous research that largely have to do with the processing of imperfect edge data, the recognition and interpretation of obscuration, and the interpretation of shadows. Mostly, the work focussed on the anomaly as a side issue, rather than as a key feature of the scene analysis.

Some of the best research on working with incomplete edge data was Falk's INTERPRET program ([Fal72]). The goal of this work was recognition of objects and structures in blocks world scenes. In the research, he recognized that edges or corners of objects are potentially missing (partially or wholely) because of obscuration or failures of segmentation. Further, he concluded that some corners could be improperly labeled because of coincidental alignments. He then attempted to sort out the problems using knowledge about edges (continuity) and of blocks (connection of edges at corners). Edges and junctions were labeled according to the topology of visible edges connecting at the junctions. With bad labelings, the type of label was used to help segment the scene into bodies. The program didn't attempt to explain the causes of the problem, only to account for a few and correct them. Further it was largely limited to corrections in the context of blocks world scenes (except for edge completions). More general scenes would not have the highly constraining vertex labelling semantics, which limits the direct use of his techniques.

When it comes to obscuration and self-obscuration, perhaps **ACRONYM** ([Bro81b]) illustrates the most thought. This program interprets scenes based on constraining the parameters of the camera and scene objects. This occurred through a matching of data to models, as mediated by a constraint manipulation system. It used a hierarchical object model based around generalized cones as primatives, that were attached using coordinate frame transforms. The model sizes and attachment parameters could be either fixed constants or variables. Image primatives were chosen to be ellipses and ribbons (2D projections of generalized cones). At any stage of its processing, it attempted to predict what objects should look like in the image, given its current value ranges for the model parameters. The prediction was a graph structure with nodes representing image structures, and arcs representing relationships that must hold between the structures. Once an image structure was matched to a predicted structure, then image size or orientation measurements were used to back-constrain the original parameter measurements. This was useful for both obtaining consistent object interpretations and deducing camera position parameters.

Brooks argues that by using geometrical reasoning, based on estimates of object position and camera parameters, **ACRONYM** can deduce that certain features are not visible (prediction of invariant and quasi-invariant features). In particular, it should be able to deduce that certain components are not visible because of obscuration or self-obscuration.

ACRONYM's geometric reasoning and image recognition have not yet been adequately demonstrated for 3D objects. The most successful examples show identification of parts for several airplanes, including sub-class identification and estimates of the camera parameters. However, as the airport scene was viewed from a large vertical height, the scene is largely two-dimensional. Further, as only a few parts are detected, the analysis seems inadequate. Brooks places much of the blame for this on poor segmentation.

Bolles ([Bol80]) considered the problem of practically recognizing two dimensional objects that may be partially obscured. His approach started with offline selection of sets of "focus features" (i.e. prominent distinguishable features, such as holes or boundaries), and discrimination tests, to allow unique identification of the object and its orientation at runtime. The identification was based on identifying focus features by their characteristics (i.e. shape and size), and the configuration of features in their neighborhood. Once a feature was uniquely identified and the configuration matched to a given model, then the object and its orientation was known. Obviously, the method required that sufficient feature information be visible.

Adler ([Adl75]), among others, used the "tee" cue for detecting obscuration. Moreover, he also applied a more sophisticated pairing heuristic, that reasoned about continuations of curved edge segments, to handle scenes like that shown in figure 2. For doing the recognition, he argues ([Adl75]) that, as "the shape characteristics of that part of the object which remains visible may be dramatically altered", one should "reduce the importance of precise shape description by exploiting context information provided by the use of models". This author's work ([Fis82] used this principle to help simplify the recognition of a table.

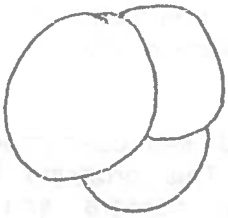

Figure 2 - complex of three obscuring regions

Waltz ([Wal75]) considered several of the anomalies in his celebrated work on segmenting blocks world scenes. The major task involved selecting a set of candidate labels for various line junctions, and algorithms for finding a compatible labelling for the scene from these labels. A prominent feature of his scenes were shadows cast by the blocks. His junction labellings explicitly included shadow constructions. Moreover, he pointed out that shadow information is needed to determine some contact and support relations. Further, he included some junction types caused by common missing line and coincidental alignment configurations. Though this work was limited to the highly constrained blocks world and required essentially perfect line/vertex data, it suggests some of the possible semantics for real scenes.

Yin ([Yin81]) reported work on recognizing overlapping two-dimensional objects. His approach was based on matching angles on the boundary of a region to those of the models. Various constraints were used to reduce the search space. Though the actual matching heuristic is probably limited to straight edge segments, the matching process is more relevant. This involved trying to maximally explain the data, assuming that obscuration would prevent perfect data from being available. Perkins ([Per77]) used a similar approach in recognizing two dimensional automobile parts. His work also covered real objects, curved segments, missing data and objects partially in the field of view.

Several people have used special techniques to find faint edges, such as in the presence of low contrast (Shirai) or high noise (Yachida). Shirai ([Shi75]) used an edge proposing approach for suggesting where to look, and a low level peak finder for suggesting the most likely edge position along a cross section to the hypothesized line of tracking. Yachida ([Yac79]), used a combination of optimization, based on models of the expected characteristics of the edge, and search among the alternative hypotheses for the continuation of an edge. These techniques are useful for the verification task, when models predict the presence of an edge, yet none are found, and secondary evidence suggests that the edge should be visible. It is noted that the edge may not be visible, as when obscured or in shadows.

Freuder ([Fre77]) looked at the problem of recognition from the viewpoint of active reasoning in a heterarchical programming structure. His SEER program was designed to reason about current "particular knowledge" (i.e. found structures) and "general knowledge" (i.e. object models) to decide what to do next. The models combine properties of subcomponent regions and relationships that hold between them. Instances of structures recognized provide "suggestions" (using the data structure/model) about other structures to look for, and any structural relationships provide "advice" about where to look. The sequence of reasoning is controlled by a priority ordered queue of suggestions. While this method of active reasoning is probably not appropriate for this project, the general approach is directly relevant.

5. Supporting Research

There are several areas of research that contribute towards the complete thesis project, but not specifically to the topic of coping with anomalies.

The first major area is philosophical support. Recent work in vision has started to concentrate on what are the relevant data, symbolic descriptions and relationships (i.e., semantics) at the various levels of image understanding. The importance of this new style of research is that it attempts to make explicit the data available at the different levels of representation, and the constraints underlying the processes transforming data between the different levels. The research proposed here is concentrated on the semantics of the intermediate level of scene interpretation; that is, on the relationships between those features whose interpretation is independent of specific objects (e.g. boundaries, surfaces, regions of special interest) and those features of specific objects themselves. Below, we briefly summarize some of the important work done in this style.

At the low level, there is good specific examples in the work of Horn ([Hor75]) on surface shape, Woodham ([Woo77]) and Brooks ([Bro81a]) on surface orientation, and Beattie ([Bea82]) on the semantics of boundaries during the image forming process. Barrow and Tenenbaum ([Bar78]) considers the low level in general in their discussion on the various intrinsic images (reflectance, illumination, orientation, distance), and then some of the semantics of boundaries in scenes.

At the intermediate level is some recent work by Lowe and Binford ([Low81]) on interpreting shadows and boundaries as evidence of three-dimensional structure. The work of Shirai ([Shi75]) on segmenting blocks world scenes is also relevant. He uses the semantics of rectangular planar surface objects (i.e. the types of junctions) to suggest where to look for the boundaries separating the planes. This is a form of knowledge based interpretation in which the knowledge used is conceptually close to the task. Binford ([Bin81]) takes a direct approach to the intermediate level semantics in his work on inferring surfaces. He uses three sources of information: depth ordering cues, relative orientations of regions from relative shadow orientations, and region boundary junction interpretations. A second important point he makes is on the preferential ordering of hypotheses, according to scene likelihoods, when interpreting ambiguous phenomena.

This author is uncertain about what constitutes good semantics at the highest levels of interpretation (i.e., "What is this scene?"), but perhaps Tenenbaum's ([Ten73]) scene based constraints are early examples (e.g. "Door is hinged to wall on one vertical edge and adjacent to floor on bottom").

The second area of related research is that of body modelling. These are methods for representing three dimensional objects in a computer system, for the effective performance of some task. Cameron ([Cam82]) reviewed and critiqued several body modelling systems from the point of view of robotics applications. Brooks ([Bro81b]) described the approach used in ACRONYM, with generalized cones as the primitives in a subcomponent hierarchy. Agin ([Agi79]) reported on a generalized cylinder based modelling system (roughly equivalent to generalized cones). The system supports symbolic parameters (e.g. class type models) and hierarchical descriptions, with transformations and attachment points. This work appears to be somewhat less advanced than the modelling used by ACRONYM, but is reported in greater detail. The function of the latter modelling systems is to support structures that can be recognized in a model-based recognition system. This research will need models that support this activity as well as provide a basis for reasoning about anomalies. The desired features include location of boundaries, location of opaque surfaces and some form of object class mechanism (as compared to just models of specific objects).

The final area is that of model-based object recognition. At one extreme there is Barrow and Burstall ([Bar74]), who looked for objects as instances of maximal cliques in a combined object model - observation graph. It is this author's opinion that this method is too abstract to effectively account for the complexities of real objects, described by real data, in real scenes. These complexities include the anomalies listed previously, such as obscuration and incomplete segments. At the other extreme is ACRONYM ([Bro81b]), which operates over given model structures, predicts an observability graph, and incrementally constrains the possibilities on objects, orientations and camera positions, via a constraint manipulation system. Other object recognition experiments include Barrow and Popplestone ([Bar71]), who investigated recognition of more realistic objects, such as teacups and eyeglasses. They used a region based segmentation, and then constructed a graph of the relations between the regions. Recognition was based on finding the prototype object whose region relationships were "closest" to that of the unknown object. Shirai ([Shi78]) also used a real object domain, and recognized telephones, desk lamps, etc. His analysis was based on edge descriptions of the scene. The edges were fitted to curve models, which were then used to suggest possible objects. Further local curves were used to substantiate the hypothesized match. Perkins ([Per77]) has developed a system for recognizing industrial parts, that, though limited to largely two-dimensional parts, copes with parts overlap and imperfect data, such as the object not being completely in the field of view. Adler ([Adl75]) used models of curved regions as part of a cartoon figure analysis program, and includes explicit analysis of obscuration by two-dimensional curved objects. Vamos et. al. ([Vam79]) demonstrated recognition of a three-dimensional industrial part, using a syntactic/fuzzy matcher. They also estimated the image transformation needed to create the image from the model.

6. More Detailed Plans

This section overviews what is felt to be the major work needed to do this project. This includes the preparatory work, the innovations and the experimental evaluation of the results.

The preparatory phase covers two areas of infrastructure: the selection and implementation of a body modeller and low level segmentation processes. How much work will be required is a serious uncertainity.

Based on the previous discussion of anomalies, it seems that what is needed in a body model is a notion of the boundaries of the object (convex extrema) and the surfaces between the boundaries. This presents a problem when it comes to curved surfaces - which have continuous extrema. The thesis may be limited to largely planar surface objects, depending on the rate of progress. A consideration is the type of information obtainable from the segmentation processes (see below). At present, it is felt that the edges can be represented by connected spline based segments, and the surfaces by elliptical surface patches. Both of these representations have reasonable properties under perspective transformations. For the model itself, it is not yet clear whether it will involve hierarchically structured subcomponents (as in **ACRONYM** ([Bro81b])) or have a uniform level of representation. One factor affecting the choice of model is the author's belief that much of object recognition must be done without the aid of a constraint maintenance and reasoning system (e.g. **ACRONYM**) and hence the modelling should be based more on logical and structural relations than on metrical ones. Yet, in order to reason about some scene properties, rough calculation of object location and orientation will be needed. However, it can be determined if two object boundaries are aligned, or if one object obscures another, without precise metrical knowledge. In conclusion, a modeller whose primary emphasis is on surface and boundary properties, and only secondarily on position and size, is needed.

Turning now to the segmentation processes, it is felt that reasonable descriptions of the boundaries and surfaces of objects will be needed. Adequate boundaries can probably be found by using a reasonable tracking edge finder, such as Beattie's ([Bea82]), Nevatia and Babu's ([Nev79]) or MacVicar-Whelan's ([Mac81]). The boundary segments, represented by splines, can be found by either tracking or by a modified Hough transform technique ([Bal81]). The surface regions can perhaps be found by aggregation based on local similarities, such as intensity statistics or micro-texture. The best elliptical region is extracted from the data, with the residuals forming other ellipses. The goal of the segmentation is to provide a reasonably complete set of low level data without many false boundaries or regions. Correct, perfect and complete data is not expected; nor is it considered to be feasible in any segmentation process that lacks scene, surface and illumination models. The requirement on the data is that it should be sufficiently complete that the anomaly based reasoning can be applied successfully.

Again it is stressed that modelling and segmentation are not the research to be pursued, so that the effectiveness of previous research will limit the range of scenes and objects recognized. This may not limit the research of the project, rather only the demonstration of the range of its applicability. However, if an inadequate foundation is chosen, then substantial efforts may be required to overcome largely irrelevant problems.

After the preparation phase, there is the research phase, which consists of:

Detailed study of the anomalies and their effects

Elaborating the rules for reasoning about the anomalies, considering any scene or segmentation requirements, and making estimates of their completeness, effectiveness and usefulness.

Formalizing the decision criteria for the application of the rules

Designing the secondary tests (such as for weak contrast boundaries)

Designing the reasoning structure into which the rules may be effectively integrated

Testing the rules over both contrived settings and general scenes

This effort is expected to require about 1 - 1.5 years. The model selection and segmentation phase is about 4 months, the research phase about 6 months, and the final integration and evaluation phase 4 months. Of course, in practice, the phases are a bit mingled. Further, because of the nature of research, these estimates are only a guideline. Because of the time constraints of this project, not all of the anomaly based reasoning will be implemented. No priority ordering has been determined yet, though it would be reasonable to order the anomalies according to frequency of occurrence and the difficulty of reasoning.

7. Anticipated Problems

The following problem areas are anticipated:

1. input data not rich enough. To recognize an object, some percentage of the object's segments must be present and reasonably correct to give sufficient support for hypothesizing the existence of the structure and locating remaining segments. A related problem is the scene itself failing to provide sufficient clues for the program to succeed, such as when processing substantially obscured objects, or differentiating between similar objects whose distinguishing characteristics are not visible.

2. input data with too many false segments. This would affect both the efficiency and correctness of the matching. At some level, the false segments will start the hypothesizing of non-existent objects.

3. modelling not rich enough. This limits the number of object types recognizable. At present, simple planar and cylindrical surface models are anticipated to cover most of the objects visible in the laboratory scene.

4. recognition of non-rectangular objects. The IMAGINE recognition program has been shown to work on rectangular structures, such as tables ([Fis82]). How well it will work on other object classes is uncertain.

5. generalizing the anomaly rules to cover model classes. It's easy enough to make specific rules that can cover for a particular anomaly observed in a particular image. How effective they can be when generalized to reason over models and scene properties is unknown. This constitutes a major question of the thesis research.

6. control of reasoning. Even without the anomaly based reasoning, the matching process can get out of hand, because of combinatorial problems. With the introduction of reasoning based not on a limited number of segments actually present, but on potentially unlimited hypotheses about incomplete data, the problems become more serious. Specific problems are:
   a. the proliferation of hypotheses (quantity of data),
   b. how long to wait for segment-driven reasoning to occur before considering anomaly-driven reasoning (control).

c. how to tell when to not bother with the anomaly-driven reasoning because the segment-driven reasoning succeeded (logic).

It is thought that the reasoning should be divided into several phases, such as:

a. over image level properties, such as boundary continuity
b. over structural reasoning, such as collection of segments into bodies
c. over particular objects from the database and the scene configuration

There should probably be over-suggestion at the lower levels, with unreasonable hypotheses being eliminated later, instead of an impoverished set of hypotheses and more sophisticated subsequent reasoning (which is more likely to be scene and model specific). The hypothesis generation would be the upward flow of information in the reasoning process. The downward flow is the suggestions for further tests at lower levels, based on the detected anomalies.

7. implementation. The address space of the PDP 11/60 is very small relative to the task, and even without the additional processing overheads incurred in circumventing the space problem, it's really not fast enough. The effect of this is to require a fair attention to largely irrelevant programming tasks.

8. Originality of the Project

This thesis project is thought to be original in the following areas:

1. the point of view on scene interpretation - that the anomalies listed in section 2 are actually normal aspects of real scenes and images, and that any complete scene interpretation program is obliged to cope with and explain their presence in the image. This includes the subsidiary requirement of including the explanations of the reasoning behind the recognition process as a part of the output.

2. the view that the anomalies listed in section 2 are a neglected part of the semantics of scenes and images when considered at an intermediate (structural) level of interpretation. Associated with this is a study of the properties of the image formation and analysis process, from the viewpoint of anomalies, and the development of a model and acquisition of meta-knowledge of the subprocesses and their normal and anomalous behavior. This also includes the consequent listing of what the anomalies are, what their properties are, how they might be detected, and how circumvented.

3. the generalization of this meta-knowledge based on a class of models for objects and the image formation process, rather than on specific instances of objects or subcomponents.

4. the implementation of programs based on a collected and consistent set of rules reasoning about anomalies, as abstract properties of images, rather than as special cases handled by specific reasoning.

5. the implementation of a more robust scene analysis program, based

on structural analysis and supported by the anomaly based reasoning. (The implementation would be limited in its variety of recognizable objects.)

9. Revisions (July 1983)

This section summarizes how the proposal has evolved over the past year. In essence, this has been a refinement of the proposal given above, to deal with the specific case of anomalies caused by obscuring surfaces.

The major claim made is that solving the problem of obscured features requires knowing where the obscured and other features should ordinarily appear in the image. This doesn't necessarily mean that the exact numerical predictions, nor exact three-dimensional scene understanding is required; however, the approach discussed here will use a certain amount of this type of information.

The first problem is knowing when obscuration occurs. In a previous paper ([Fis83]), it was noted that obscuration (anomalies 15,16,17) caused structure loss. Hence, missing structure should be a clue to the presence of obscuring objects.

Now, this structure loss can be apparent at all levels of description - from the absence of specific objects, to the absence of specific features on the surfaces of objects. This leads to the specific problem of detecting loss of information, for each level of description.

In particular, three levels of representation are considered:

- image regions
- surfaces
- objects

The image region level is primitive, so there is never any "missing" structure. The surface level description details what features of the surface have evidence and where that evidence comes from. Finally, the object level description will contain what substructure (sub-objects or surfaces) are found and what the supporting evidence for these claims are. Hence, for the surface and object levels of description, whenever there is missing data, we may hypothesize obscuration.

The next problem is that of verifying the hypotheses. For wholly obscured structure, it may be possible to deduce when the structure will not be visible, because of it being on the back-side of an object. Further, it may be possible to predict when front-facing structures will be obscured by closer structures.

For obscuration caused by external, unexpected structure, more circumstantial evidence has to be accepted. The best evidence would be to know where the structure should be in the scene, and to show that there is structure closer to the viewer. This requires showing that all regions in the image, located where the predicted structure should be seen, correspond to closer surfaces. This can be done directly, by locating the obscuring objects in space, or indirectly, by finding cues (such as Tee junctions), that support the ordering relationship between the hypothesized object and the obscuring surfaces.

For partially obscured objects, the problem can be ultimately reduced to the surfaces that make up the object. The surfaces may be complete, which presents no problem, totally obscured, which was discussed above, or partially obscured. With partially obscured surfaces, the approach would be to show that what is visible is consistent with the object and that any non-object boundaries are consistent with being obscuring boundaries. Evidence that supports this is similar to that of the wholly obscured case.

Several requirements have to be met to do the hypothesis formation and verification process. First, we need to have a representation of the objects and surfaces, so that we can tell when data is missing. Secondly, there must be a process that instantiates the representation, even in the face of incomplete data. Lastly, a procedure for verifying the fully instantiated hypotheses is needed.

For the first requirement, a surface-oriented body modeller will be used for the objects. This will make explicit the major external surfaces of the object, along with their relative geometrical relationships. The surfaces will be described by their boundary shapes. These assumptions limit the classes of objects to planed or simply curved surfaces, with seams. Further, having a geometrical model necessitates a certain amount of geometrical reasoning.

For the second requirement, several components will be needed. These subcomponents are grouped into several sets:

- subcomponents that make surface hypotheses from one or more image regions,
- subcomponents that match model surfaces to "whole" surface hypotheses, simultaneously extracting the three-dimensional mapping parameters,
- subcomponents that integrate these matches to form nearly-complete structures, and
- subcomponents that fill in the missing bits, based on various forms of obscuration-based (or other) reasoning.

The third requirement is for a verification process. The hypothesizing process is based on clues suggesting the match between image features and models. The verification process will ensure that the final hypothesized structure is consistent with the data and with known properties of structures. In particular, this includes logical properties, such as:

- image features can be associated with only one model feature, and
- adjacent model features must have adjacent image features.

Also needed are some tests to ensure that structures hypothesized by the obscured surface reasoning and evidence structure are both totally consistent, and together make up the complete object. Lastly, it would be desirable to declare that the hypothesis structure is fully instantiated.

The novelty of this refined proposal still generally lies under the categories discussed in section 8. There are, however, three specific technical innovations that the project features. These are:

- The processing will make explicit use of surfaces as a hypothesized entity, and will have specific rules for hypothesizing surfaces from image features.
- There will be a process that matches surface hypotheses to model surfaces, with the goal of making both good matches, and extracting the six positional parameters of the match. This process will be based on both object shape and input surface data.
- Obscured features will be expected and handled with at both the

levels of surfaces and objects.

For each of these three areas, various amounts of theoretical work have been done. The importance of explicit surface hypotheses is highlighted by Marr in his discussion of the 2 1/2 D sketch ([Mar82]). Stevens ([Ste81]) has demonstrated one form of surface shape to orientation understanding. (The proposed work will relate some of his results to the matching of surface shapes to specific model surfaces.) Binford ([Bin81]) discusses how to extract three-dimensional information from image features; in particular, many of the ideas on surfaces and obscuration follow from his ideas. So, in some ways, this proposal is for a largely "experimental" thesis project, that implements the general theoretical ideas and validates them in a specific form.

10. Bibliography

[Adl75] Adler, M., **"Understanding Peanuts Cartoons"**, Edinburgh Research Report #13, 1975.

[Agi79] Agin, G.J., **"Hierarchical Representation of Three-Dimensional Objects Using Verbal Models"**, SRI Technical Note #182, March 1979.

[Bal81] Ballard, D.H., **"Generalizing the Hough Transform to Detect Arbitrary Shapes"**, Pattern Recognition, Vol 13, #2, pp111-122, 1981.

[Bar71] Barrow, H.G., Popplestone, R.J., **"Relational Descriptions in Picture Processing"**, in Meltzer and Michie (eds), Machine Intelligence 6, pp377-396, 1971.

[Bar74] Barrow, H.G., Burstall, R.M., **"Subgraph Isomorphism, Matching Relational Structures and Maximal Cliques"**, Edinburgh Dept. of Artif. Intel. Working Paper #5, 1974.

[Bar78] Barrow, H.G., Tenenbaum, J.M., **"Recovering Intrinsic Scene Characteristics From Images"**, in Hanson & Riseman (eds), Computer Vision Systems, pp3-26, 1978.

[Bea82] Beattie, R.J., **"Edge Detection For Semantically Based Early Visual Processing"**, Proc. 1982 European Conf. on Artif. Intel., forthcoming.

[Bin81] Binford, T.O., **"Inferring Surfaces from Images"**, Artificial Intelligence, Vol. 17, pp205-244, 1981.

[Bol80] Bolles, R.C., **"Locating Partially Visible Objects: The Local Feature Focus Method"**, SRI Technical Note 223, Aug 1980.

[Bro81a] Brooks, M.J., **"Surface Normals From Closed Paths"**, Proc. 6th IJCAI, pp98-101, 1979.

[Bro81b] Brooks, R.A., **"Symbolic Reasoning Among 3-D Models and 2-D Images"**, Artificial Intelligence, Vol 17, pp285-348, 1981.

[Cam82] Cameron, S. **"Body Models for Every Body"**, Edinburgh Dept. of Artif. Intel. Working Paper #107, 1982

[Fal72] Falk, G., **"Interpretation Of Imperfect Line Data As A Three-Dimensional Scene"**, Artificial Intelligence, Vol 3, pp101-144, 1972.

[Fis82] Fisher, R.B., **"A Structured Pattern Matching Approach to Intermediate Level Vision"**, Edinburgh Dept. of AI Research Paper #177, 1982.

[Fis83] Fisher, R.B., **"Anomalies Occurring in the Image Formation and Analysis Process"**, Edinburgh Dept. of AI Working Paper #137, 1982.

[Fre77] Freuder, E.C., **"A Computer System For Visual Recognition Using Active Knowledge"**, Proc. 5th IJCAI, pp671-677, 1977.

[Hor75] Horn, B.K.P., **"Obtaining Shape from Shading Information"**, in Winston (ed), The Psychology of Computer Vision, McGraw-Hill, New York, 1975.

[Low81] Lowe, D.G., Binford, T.O., **"The Interpretation of Three-Dimensional Structure from Image Curves"**, Proc. 7th IJCAI, pp613-618, 1981.

[Mac81] MacVicar-Whelan, P.J., Binford, T.O., **"Intensity Discontinuity Location to Subpixel Precision"**, Proc. 7th IJCAI, pp752-754, 1981

[Mar82] Marr, D., Vision, W.H. Freeman & Co., 1982.

[Nev79] Nevatia, R., Babu, K.R., **"Linear Feature Extraction and Description"**, Proc. 6th IJCAI, pp639-642, 1979

[Oht81] Ohta, Y. Maenobu, K., Sakai, T., **"Obtaining Surface Orientation from Texels Under Perspective Projection"**, Proc. 7th IJCAI, pp746-751, 1981.

[Per77] Perkins, W.A., **"Model-Based Vision System for Scenes Containing Multiple Parts"**, Proc.

5th IJCAI, pp678-684, 1977.

[Shi75] Shirai, Y., **"Analyzing Intensity Arrays Using Knowledge About Scenes"**, in Winston (ed), The Psychology of Computer Vision, ch 3, pp93-113, 1975.

[Shi78] Shirai, Y., **"Recognition of Real World Objects Using Edge Cue"**, in Hanson and Riseman (eds), Computer Vision Systems, pp353-362, 1978.

[Ste81] Stevens, K.A., **"The Visual Interpretation of Surface Contours"**, in Brady (ed.), Computer Vision, North-Holland, 1981.

[Ten73] Tenenbaum, J., **"On Locating Objects By Their Distinguishing Features In Multi-sensory Images"**, SRI Technical Note #84, Sept 1973.

[Vam79] Vamos,T., Bathor,M., Mero,L., **"A Knowledge-Based Interactive Robot-Vision System"**, Proc. 6th IJCAI, pp920-922, 1979.

[Wal75] Waltz, D., **"Understanding Line Drawings of Scenes with Shadows"**, in Winston, The Psychology of Computer Vision, pp19-91, 1975.

[Win75] Winograd, T., **"Frame Representations and the Declarative/Procedural Controversy"**, in Bobrow & Collins (eds), Representation & Understanding, pp 185-210, ch. 7, 1975.

[Wit80] Witkin, A. **"A Statistical Technique for Recovering Surface Orientation from Texture in Natural Imagery"**, Proc. 1st AAAI Conf., 1980.

[Woo77] Woodham, R., **"A Cooperative Algorithm for Determining Surface Orientation from a Single View"**, Proc. 5th IJCAI, 1977.

[Yac79] Yachida, M., Ikeda, M., Tsuji, S., **"A Knowledge Directed Line Finder For Analysis of Complex Scenes"**, Proc. 6th IJCAI, pp984-991, 1979.

[Yin82] Yin, B.L., **"A Program Which Recognizes Overlapping Objects"**, Edinburgh Dept. of AI working paper #93, 1981.