June 1989

Research on Symbolic Inference in Computational Vision

Ruzena Bajcsy  
*University of Pennsylvania*

Alok Gupta  
*University of Pennsylvania*

Max Mintz  
*University of Pennsylvania*

Follow this and additional works at: [http://repository.upenn.edu/cis_reports](http://repository.upenn.edu/cis_reports)

Recommended Citation  


This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/cis_reports/839](http://repository.upenn.edu/cis_reports/839)  
For more information, please contact repository@pobox.upenn.edu.
Research on Symbolic Inference in Computational Vision

Abstract
This paper provides an overview of ongoing research in the GRASP laboratory which focuses on the general problem of symbolic inference in computational vision. In this report we describe a conceptual framework for this research, and describe our current research programs in the component areas which support this work.

Comments

This technical report is available at ScholarlyCommons: http://repository.upenn.edu/cis_reports/839
RESEARCH ON SYMBOLIC
INFEERENCE IN
COMPUTATIONAL VISION

Ruzena Bajcsy,
Alok Gupta
and Max Mintz

MS-CIS-89-41
GRASP LAB 183

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104

June 1989

Research on Symbolic Inference
In Computational Vision

Ruzena Bajcsy, Alok Gupta, and Max Mintz *
Department of Computer and Information Science
GRASP Laboratory
University of Pennsylvania
Philadelphia, PA 19104-6389

5 May 1989

Abstract
This paper provides an overview of ongoing research in the GRASP laboratory which focuses on the general problem of symbolic inference in computational vision. In this report we describe a conceptual framework for this research, and describe our current research programs in the component areas which support this work.

1 Introduction

This paper provides an overview of ongoing research in the GRASP Laboratory which focuses on the general problem of symbolic inference in computational vision. In this overview we describe a conceptual framework for this research, and describe our current research programs in the component areas which support this work.

Historically, research in computer vision has addressed the general problem of attaching symbolic labels to image regions based on selected transformations of the observed (vision sensory) data. Pentland (1986) refers to this process as transformations from pixels to predicates. Symbolic representations have played a central role in AI from its very beginnings as is evident from the early work of McCarthy, Minsky, Newell, and Simon. A critical aspect of symbolic representations lie in their inherent compactness, i.e., the exchange of a generally high-dimensional data set for a compact symbolic label. This important compactification is achieved at a price – namely a loss of information and the attendant approximation which is inherent in the selection of a symbolic label. The underlying approximation process is akin to the selection of a partition of a set, which, in turn, determines an equivalence relation. The symbolic label denotes the appropriate equivalence class within the partition.

Researchers in computer vision, e.g., Winston (1984), have also considered the inverse problem, i.e., the transformation from symbolic labels to predictions of pixel-level (signal) data. Here, the problem becomes: Given high-level knowledge of an object or objects in the field of view, what are the predicted or expected sensory data for given sensor positions, geometry, etc? In this setting, the symbolic label represents the object of inquiry. This top-down analysis provides specific procedures to be applied to the sensory data for the purpose of extracting the expected features associated with the symbolic representation.

The signal to symbol and symbol to signal transformations both rely heavily on underlying models. These models include physical and mathematical models such as: object material properties (reflectance, smoothness, texture), light sources (illumination, geometry), geometric primitives (manifolds, surface patches, shape models), sensor models (sensor geometry, resolution, error models), and models of a priori information (object classes, object-class variability, parametric bounds). In addition to these models, there are task-dependent models which implicitly or explicitly set upper and lower limits on the quantization (partition) errors associated with the selection of symbolic labels for the objects of inquiry and component high-level features. The signal to symbol and symbol to signal transformations and interrelations are depicted in Figure 1.
In realistic computer vision applications one observes that the signal to symbol transformations and the symbol to signal transformations lead, respectively, to errors in the inferred symbols and errors in the estimates of the desired parameters and features. This error behavior can, in principle, be improved by using feedback strategies. The concept of feedback mechanisms to improve control system behavior is of ancient origin. In this setting, the feedback processes take two forms: In the inference process for symbol selection, one is inherently limited to a finite set of symbols. Further, the decision models can be structured using finite-graph models. Thus, the error checking and error correction processes for symbolic inference can be viewed as graph searching processes. These graph searches take the form of backtracking. Whereas, the processes for estimating expected image features based on object models are primarily processes with continuous decision spaces. The estimation processes are based on well-established physical and mathematical relations which constrain the estimation processes. These feedback paradigms are depicted in Figure 2.

Figure 1.

Signal --> Symbol

Symbols:

Object Classes
   /\          Object Models
   |           |
   \|/         \|/
3-D Features <---------- 3-D Models
   /\                  |
   |                  \|/
2-D Features <---------- 2-D Models
   /\                  |
   |                  \|/
1-D Features <---------- 1-D Models
   /\                  |
   |                  \|/
Image               Predicted
Data                Image Features
Figure 2.

2 Signals To Symbols — Symbolic Inference

2.1 Related Research

Various aspects of model-based symbolic inference in computational vision have been addressed by Grimson, and Lozano-Pérez (1984), Binford, Levitt, and Mann (1987), and Mulder (1988).

The paper by Grimson, and Lozano-Pérez (1984) describes how sparse measurements of 3-dimensional position and surface normals may be employed to identify and locate objects from among a set of known polyhedra which can have up to six degrees of freedom relative to the sensors. The methods in this paper do not require highly accurate measurements of the surface normals. However, the success of these methods can be critically dependent on the accuracy of the a priori models of the individual polyhedra.

The paper by Binford, Levitt, and Mann (1987) describes a model-based hierarchical Bayesian approach to symbolic inference. The system design includes comprehensive physical representations of objects and object-relations. The probabilistic updating is achieved using a Bayesian network. The goal of the system is to reason about geometry in machine vision — where there is extensive domain-dependent a priori information. The paper addresses the important issues of: (i) probabilistic sensor models, i.e., the statistical relations between model parameters and noisy sensor observations; (ii) the role of quasi-invariants and probabilities; and (iii) the role of domain information in the estimation of prior probabilities for the uncertain parameters. Although the requirement for extensive domain knowledge limits the application of the underlying system, there are many ideas in this paper which are relevant to other, less restrictive, applications.

The paper by Mulder (1988) describes a model-based vision system which is designed to deal with a large number of scene interpretations which are hypothetical and ambiguous. The
The paper defines and exploits *discrimination graphs* as replacements for specialization hierarchies. Discrimination graphs are invoked to contend with the following problems associated with specialization hierarchies: (i) discrimination graphs provide uniform representation, whereas, specialization hierarchies may provide non-uniform representations; and (ii) discrimination graphs avoid the combinatorial explosions which are associated with specialization hierarchies. Discrimination graphs represent classes of objects with similar image features. The leaves of these graphs represent classes of elementary objects which can describe an image unambiguously. The internal nodes represent abstract classes of such objects. Instead of invoking elementary object classes directly, which is common in most model-based vision systems, each image feature is mapped into the interpretation class representing the complete set of admissible (elementary) object classes. As new constraints are discovered in the image, constraint propagation techniques are used to force each interpretation class to become more specific.

### 2.2 Domain Selection

One of our research objectives is to develop a better understanding of the role of data driven processing (bottom-up analysis) and processing based on domain-specific knowledge (top-down analysis). This includes the interrelationships between these two processes. In our current research we are particularly interested in developing methods which do not require extensive domain-dependent *a priori* information. This limitation is relevant, for example, where we seek to explore relatively unknown environments. Thus, we would not expect to have the ability to make use of rich *a priori* information of the sort required by Grimson, and Lozano-Pérez (1984), Binford, Levitt, and Mann (1987), and Mulder (1988). Example domains include: (i) a kitchen, where a detailed inventory of the utensils and appliances (in the form of precise 3-dimensional models, and potential configurations) is unavailable; and (ii) a remote natural environment which has not been previously explored. In both examples, we have to contend with imprecise class descriptions, e.g., there are many different types of spoons, and forks — where we consider shape, size, structural material, etc. Similarly, stones, rocks, and bolders are nondenumerable with respect to size and shape descriptions. Our basic approach in dealing with this type of imprecise object description is via the notion of an *uncertainty class.*

### 2.3 Salient Research Issues

We have identified the following research issues in the study of symbolic inference in computational vision:

1. The selection of a finite symbol set: This selection entails the delineation of object uncertainty classes and the selection of a suitable partitions of these classes. The sets of the
partitions are the symbolic labels. An uncertainty class $C$ is a set of objects that have some scalar- or vector-valued property that can be measured and represented by a point in a given subset of $E^n$. For example, we could consider the set of all rod-like objects with given bounds on the underlying dimensions. We partition $C$ into equivalence classes by task-driven requirements such as size categories for manipulation, sensor limitations, such as resolution limits, or resource limitations such as data storage and computational limitations.

2. The encoding of the certainty or confidence in the symbolic inference process: This encoding of a confidence measure is required for: (i) ranking decisions on symbolic labels, (ii) error detection and correction, and (iii) determining if the (local) decision process terminates or if more data is required to make the symbolic inference.

The design of the confidence measures should not be limited by Bayesian considerations, i.e., we should not require knowledge of a priori probabilities. However, we should select confidence measures which can incorporate the a priori probabilities when this information is available.

3. Error detection and correction: Error detection and correction are critical issues. Here, we believe feedback control structures play an essential role. The decision to terminate the local process or to gather more data is of central importance.

4. Combination rules: We require combination rule (a calculus for combination) to merge: sensory data from multiple sensors; preliminary local hypotheses; and potentially related symbolic labels. Where probabilistic models apply, these rules may non-Bayesian or Bayesian, depending on the existence of prior probabilities. It is noteworthy that non-Bayesian procedures may have Bayesian interpretations. A famous example is that of maximum likelihood estimation with a compact parameter space. Here, the maximum likelihood estimate can be interpreted as a maximum a posteriori estimate with respect to a uniform prior distribution. While careless applications of Bayesian models can be costly, it is important to note that there are circumstances and methods for choosing a prior distribution in light of the uncertainty in its selection. A good discussion of the robustness of Bayes procedures appears in Berger (1985).

5. Stability and convergence: A well-designed decision system for symbolic inference must be robust to uncertainties in the underlying models of the objects of inquiry and the sensory systems. The system must avoid deadlock and thrashing behavior, and have built-in data-dependent mechanisms or criteria for determining convergence.
3 Research Foci

In the remainder of this report we delineate two programs which support this research on symbolic inference in computational vision.

3.1 Part Description and Segmentation Using Contour, Surface, and Volumetric Primitives

The problem of part definition, description, and decomposition is central to shape recognition systems. The ultimate goal of segmenting range images into meaningful parts and objects has proved to be very difficult to realize, mainly due to the isolation of the segmentation problem from the issue of representation. We propose a paradigm for part description and segmentation by integration of contour, surface, and volumetric primitives. Unlike previous approaches, we use geometric properties derived from both boundary-based (surface contours and occluding contours), and primitive-based (biquadratic patches and superquadric models) representations to define and recover part-whole relationships, without a priori knowledge about the objects or the object domain. The objects shape is described at three levels of complexity, each contributing to the overall shape. Our approach can be summarized as answering the following question: Give that we have all three different modules for extracting volume, surface, and boundary properties, how should they be invoked, evaluated and integrated? Volume and boundary fitting, and surface description are performed in parallel to incorporate the best of the coarse to fine and fine to coarse segmentation strategy. The process involves feedback between the segmentor (the Control Module) and individual shape description modules. The control module evaluates the intermediate descriptions and formulates hypotheses about parts. Hypotheses are further tested by the segmentor and the descriptors. The descriptions thus obtained are independent of position, orientation, scale, domain and domain properties, and are based purely on geometric considerations. They are extremely useful for the high-level domain-dependent symbolic reasoning process, which need not (and should not) deal with tremendous amounts of data, but only with a rich description of the data in terms of primitives recovered at various levels of complexity.

This research program is the subject of Section 4 of this paper and is described in detail in Gupta (1989).

3.2 Robust Fixed Size Confidence Procedures

Fixed size confidence procedures are set-valued estimators with specified geometry and size (e.g., intervals, rectangles, circles, spheres, ellipses, ellipsoids, etc) which are designed to cover
(contain) the true value of the unknown feature (parameter value) with a given minimum probability. Roughly speaking, the larger the “size” of the confidence set, the larger the resulting minimum probability of coverage. Thus, the notion of an optimal fixed size confidence procedure, is a fixed size set-valued function of the observed data, which maximizes the minimum a priori probability of coverage. The confidence procedure concept does not require any a priori probability model for the unknown feature vectors. Further, these set-valued estimators can be obtained in the face of uncertainty in the true sampling distribution (sensor noise statistics). Confidence procedures which are designed to be resistant to sampling distribution uncertainty are referred to as robust confidence procedures. Thus, the concept of a robust confidence procedure lends itself to estimation problems which arise naturally in the context of symbolic inference in computational vision. The size (dimensions) of the confidence procedure in this application area is determined by task-dependent considerations, i.e., the required accuracy of the resulting decisions. Thus, the determination of size is made “outside” the decision-theoretic framework which characterizes the confidence procedures. Confidence procedures are useful at both the “data” and “symbolic” levels of the inference process. For example, on the signal or data level, robust fixed size confidence procedures provide an important means to achieve robust multi-sensor fusion. Further, robust confidence procedures which are designed for discrete parameter spaces (symbol spaces) provide an important means to measure the reliability of the symbolic inference process. Again, it is important to emphasize that these techniques apply outside the Bayesian framework, but are able to incorporate prior probability information when it is available.

This research program is the subject of Section 5 of this paper. The multi-sensor fusion research is described in greater detail in McKendall and Mintz (1988). The research on robust confidence procedures which are designed for discrete parameter spaces (symbol spaces) is a subject of current research. One of the most interesting outcomes of this research on symbolic parameter spaces, is the delineation of the importance of nonmonotone procedures in the development a complete theory.

4 Part Description and Segmentation Using Contour, Surface, and Volumetric Primitives

4.1 Introduction

In this section we propose a paradigm for decomposition of complex objects in range images into the constituent parts based on the shape, using contour, surface, and volumetric primitives. Decomposition into parts, units or primitives is the basis of scientific methodology. Because of
the limits on how much information we can process at a time, we have to simplify and view the world at various levels of abstraction. Many reasons have been advanced in favor of such a decomposition. A recognition-by-parts approach is not sensitive to occlusion and is extremely powerful in handling countless configurations of articulated objects. A description in terms of basic shape primitives is more efficient, parsimonious in space consumption, and facilitates structured description of the world.

In computer vision literature the partitioning of images and description of individual parts is called segmentation and shape representation. We have presented arguments in Bajcsy, Solina, and Gupta (1988) that the problem of segmentation and representation are related and have to be treated simultaneously. Solving any one of the two problems separately is very difficult. Since neither of them can be easily solved in isolation, at least not on the first try, we argue that they should interact to guide and correct each other. Hence, segmentation and shape recovery should not be studied separately. The complete visual interpretation problem is even more complex because the initial data acquisition process cannot be separated from the later segmentation and shape representation. How data acquisition can interact with the interpretation stage is investigated in computer vision under the heading of active vision (Bajcsy 1989).

Since the objects in the world are of arbitrary complexity, it is not possible to include primitives for all the different shapes as it will never be a complete set. Thus we have to make a judicious choice of primitives that have the capability of describing data at various levels (dimensions), so that description at some level is always possible and computability of primitives is assured. We propose that for obtaining a global shape description from single-viewpoint 3-D data requires addressing shape at following levels:

1. **Volumetric level**: Primitives capable of modeling parts in three dimensions are needed to describe global shape of parts.

2. **Surface level**: Surface primitives describe *internal surface boundaries* and *surface patches* which are difficult to model by volumetric primitives, but are vital source of information about recovering part structure.

3. **Occluding Contour level**: The Occluding contour encodes the 3-D shape of parts projected on the image plane.

This hierarchy of shape primitives allows one to obtain shape descriptions at volumetric, surface and occluding contour level. Since, both boundary-based and primitive-based primitives are included in our vocabulary, the representation is expressive and robust. It is clear that no one primitive will always capture all the details of shape. For example, if it is not possible to model parts with the selected volumetric primitive, an approximation at volumetric level can
be obtained, with more detailed description at surface level. Thus, completeness requirement for a general representation is satisfied by obtaining hierarchical descriptions.

The criteria for selection of shape primitives have been studied extensively by vision researchers (Brady 1983, Brady 1984, Marr 1982, Binford 1982, and Rao 1988). The shape primitives should be invariant to rotation, translation, and scale. Accessibility, defined as computability of the primitive is essential, since our goal is to recover the structure from the input. Stability of the primitive with respect to minor changes due to noise or viewpoint, with respect to scale and configuration is important to generate consistent representations. While small changes in scale should not create major changes in description, a multi-scale representation should be possible, for example, parts become detail as the scale is increased. The primitives should have local support, so that occluded parts can still be described and recognized when matching is performed against stored descriptions.

Low level models like contours and edges have low granularity and are too local to capture or make use of the gross structure of the world. They are sensitive to local changes and difficult to put together in a global context. However, this characteristic allows them to capture local details of shape that would be missed or smoothed out by more global primitives. When analyzed as a whole, contour primitives have the remarkable capability of describing global shape and segmenting planar shapes into parts. Points of interest include positive curvature maxima, negative curvature minima, and inflection points on the contour.

The next level of shape description is achieved by describing local and overall surface characteristics. Surfaces play important role in human perception of shape. A lot of effort in computer vision has been spent on describing complex surfaces as piecewise continuous patches. In order to arrive at a global interpretation, a surface representation scheme that combines relevant surface contours with the surface patches is needed. Concave tangent discontinuities ($C_1$ type) in the occluding contour and surfaces provide partitioning rules for them. Zero-crossing contours on surfaces partition smooth surfaces into piecewise-continuous quadric patches. Reliable computation of these discontinuities is still an open problem in computer vision.

Three dimensional primitives like generalized cylinders and cones, polyhedral models, 3-D Smoothed local symmetries Brady (1983), and 3-D symmetric axis transform (Nackman 1985) have been used by model based vision systems. However, the power of representation varies from model to model. A model allowing deformations is likely to describe objects with fewer primitives than a rigid model which will need more instances to approximate the object. Volumetric primitives are essential to generate compact object-centered descriptions and to define global part-structure. Superquadric models, our choice of volumetric primitives, provide object centered descriptions, thus allowing surface and contour level descriptions to attach to the local coordinate system, facilitating ease in representation and model-based matching.
The problem then is how to use the primitives to segment the objects into part-structure. In the context of shape recognition, the problem of segmentation can be defined as matching the right kind of shape model with the right parts of data in an image. This brings up the crucial question of facilitating this matching process.

Each of the shape primitive can independently describe the data. The occluding contour-based segmentation is widely studied in pattern recognition and computer vision as 2-D shape recognition problem (Pavlidis 1977, Shapiro 1980, and Asada and Brady 1986). Surface based approaches have been popular with model-based vision systems, as they have local support, and allow 3-D objects to be modeled as collection of surfaces. Volumetric models have proved to be most difficult to recover from image data. Some researchers have used a combination of features to model domain specific objects (Brooks 1983), exploiting the robustness achieved by combining descriptions at different levels. To facilitate segmentation we believe that for a general purpose vision system one needs volumetric, surface and boundary shape primitives. Difficulty in recovering volumetric models in intensity images is experienced due to the loss of depth information. But the problem has not proved to be any easier even with the availability of depth information (Nevatia and Binford 1977, Solina 1987, Boult and Gross 1987, Rao 1988, and Soroka and Bajcsy 1978). We are considering the input to be dense depth maps, scanned by an active range scanner from a single viewpoint. No information about scanner geometry or viewpoint is required.

4.2 Research Proposal: An Integrated Approach

Having described the shape primitives and identified the role of each primitive in shape segmentation and description, we now focus our attention on the goal of this research, which is to develop an effective control structure that works in conjunction with these modules to extract the part-structure of a complex object. The primitives give a hierarchy of shape descriptions, ranging from the planar contour level to the three-dimensional volumetric level. The problem that we wish to solve can be stated in the following way. Given that we have all three different modules for extracting volume, surface and boundary properties, how should they be invoked, evaluated and integrated? There are two possibilities. The first one is to apply all three modules simultaneously. The second is to apply them strictly in a predetermined sequence. In the parallel approach conflicting hypotheses can arise that would have to be resolved. The sequential method may lead the segmentation process in a wrong direction so that backtracking would sometimes be necessary. A combined approach where all three methods could interact would not be so vulnerable. This opens up the problem of evaluating and comparing information embedded in models built by different aggregation methods. How to evaluate the models indi-
4.2.1 Motivation

Before we propose our control strategy, it is instructive to study the behavior of the shape primitives on the actual data consisting of objects of varying complexity. The volumetric shape recovery procedure (Solina 1987) was applied to a set of range images of single objects (Figures 1 to 6). The contour obtained by tracking the occluding boundary and the contour of the recovered volumetric model are compared in all the cases. For the objects in figures 4 to 6, surfaces reconstructed from the superquadric model are compared with the original range data.

While the volumetric model gives a holistic explanation of the whole object it can miss details that are beyond the scope of the model. An overall measure of goodness of fit, like the residual from least-squares fit, or the distance measure does not always give an accurate evaluation of the appropriateness of the volumetric model. Although models can have acceptable overall goodness-of-fit, like the volumetric model for the box with cut-out (figure 1), they need not be the acceptable representations of the object. On the other hand, for value of the goodness-of-fit in same range, volumetric models for the vase (figure 5) and the box-with-jagged-edge are more
Figure 3: A composite object (cylinder glued to box): The poor approximation of the object reflects need for segmentation.

Figure 4: Object with parts (a wrench): The two boundaries coincide in only part of the image alerting to the fact that the object has parts.

or less acceptable volumetric representations of the actual object. This argues for a measure other than the quantitative measure of goodness-of-fit or Euclidean distance. The qualitative measure obtained by comparing the local boundary of the object in the range image with the boundary of the recovered volumetric model can point out the limitations of the volumetric model and suggest improvements in segmentation or refinement in shape representation. When boundaries do not coincide, preference should be given to actual boundary in the range image, but the possibility of missing data (due to self occlusion) must also be considered.

The Part versus detail issue can be addressed at individual primitive levels as well as collectively. For example, the vase in figure 5 is formed of three second-order surface patches, collectively organized in a cylindrical shape. At the volumetric level, a cylindrical model is sufficient to describe the overall shape. Details have to be obtained in terms of second order
Figure 5: **Object with surface detail (A vase)**: The difference between the two outlines is negligible compared to the overall size of the object. However, to recover more detail, and to define the internal boundaries, surface description is necessary.

Figure 6: **Object with hole and cavity**: Surface and contour information is required to effectively segment it into parts and to define concavities on the surface.
patches at the surface level. Contour analysis signals the presence of details on the object, and accepts the superquadric model. However, the superquadric model is accepted only after the surface comparison yields acceptable error. Thus, both the qualitative measures are essential for model evaluation. The presence of details in the form of a jagged edge is similarly detected in figure 2. It should be noted that the details are not neglected in the final description. They are ignored by only the volumetric model. Contour and surface description are generated in detail with the final decision of assigning labels postponed to the domain-dependent processing. For example, a pitcher's small dent on the rim is necessary for recognition, so it cannot be ignored by a bottom-up shape description process. However, the decision to segment the object into volumetric primitives has to be taken at the geometric level.

Closely tied to the issue of part-detail is the issue of part-whole relationships. What cannot be brushed away as a detail has to be considered a part at the volumetric level. It is easy to detect presence of distinct parts in the object (figures 3, 4 and 6), by contour and surface comparisons. It is another matter to recover them in terms of primitives. It needs partitioning the object into parts at surface boundaries and contour concavities. How do surfaces and contours interact to generate hypotheses about parts and then use superquadrics to verify the hypotheses? What if there is no volumetric description possible for the part? What is the best approximation for such a part? What do we mean by acceptable shape description? To attempt answers to these questions we propose our approach next.

4.2.2 The Proposed Approach

The detailed flow diagram of our proposed approach is shown in figure 7. The past research of 3-D part segmentation has been mostly theoretical. To satisfy the practical constraints of computability and robustness we propose a parallel closed-loop segmentation process with active feedback between different description modules. From the examples in the previous section it is clear that interaction among different primitives is imperative.

To incorporate the best of the coarse to fine and fine to coarse segmentation strategy we propose to perform volume, surface, and boundary fitting in parallel on the input data. The volumetric shape recovery is a global method, going from very coarse to fine fitting on the part level while surface and boundary detection going from fine to coarse. These two processes are complementary in the approach of explaining the data, accounting for global position, orientation, size and shape such that the descriptions obtained at the global and local levels support each other. Thus, it is the local processing by the Occluding contour and the Surface modules that is done in parallel and has to be done only once. The global description at the contour and surface level is obtained by refining these initial measures in a closed-loop
feedback. The Curve Segmentation module and the Surface Segmentation module perform the refinements in a typical fine to coarse manner through an internal feedback as well as an external feedback from the control module (figure 7). For example, fitting global second order patches on the surface needs intra-primitive feedback from the surface level itself, while detecting surface boundaries also needs inter-primitive feedback from the occluding contour. The segmented descriptions are evaluated and integrated at the inter-primitive level by the control module along with the evaluation of superquadric model to combine the descriptions. Since the superquadric model estimation treats data globally, the initial estimation might not be acceptable due to presence of parts. Once the control module (the global segmentor) generates hypotheses about parts, the superquadric procedure gives the best fitting models for verification of the hypotheses. Thus the model recovery procedure works as the hypotheses verifier at the volumetric level. It then follows that part-segmentation is the core of the problem.

To achieve an effective segmentation of a single viewpoint scene, the control structure has to determine the reliability of information obtained from each primitive. Superquadrics being part-models, need to be compared with the bounding contour and available surface points to evaluate suitability of the recovered model. Surfaces, for most part, complement the information provided by bounding contours. Bounding contours are viewpoint dependent and may
not account for all relevant contours needed for complete segmentation or description. This is obviously the case when viewpoint is not general. Thus, in some cases, when volumetric information is not available, surface information along with bounding contour can determine if the object is in a general position or not and ask for information from different viewpoint (or rotate the object). For some objects, it may not be possible to obtain data from a viewpoint such that the object can be segmented by analyzing only the contour. In such a case, if surface information strongly suggests segmentation along a surface discontinuity, bounding contour should not lower our confidence in surface information. On the other hand, if contour suggests a possible segmentation and there is no support from surfaces, a decision will have to be made about the possibility of segmentation assuming a possible smooth join between part and object body. Superquadrics essentially provide global description of individual parts and give the feedback as to the possibility of a further segmentation of that part. They lack the local information needed to suggest possible segmentation sites. Contour and Surfaces, on the other hand, actively hypothesize and carry out segmentation. The process continues until a satisfactory description of parts is achieved.

How do we evaluate the intermediate descriptions? As seen in the examples, the global feedback loop between the individual descriptors and the control module gives a set of "difference measures" at the contour and surface level. Many techniques are available for planar contour matching and surface matching in pattern recognition literature. We want to use this feedback for evaluation of the intermediate descriptions as well as for further segmentation. The differences can be interpreted as "overestimation" or "underestimation" of actual data by recovered models. Since superquadrics tend to undersegment (figure 3), and bring in symmetry considerations, the difference patterns generated by them consist of overestimated and underestimated regions (e.g. cup in figure 6).

What do you do if different types of models do not mutually reinforce each other? In such cases, one would normally prefer models of smaller granularity that are less prescriptive models that closely follow the data in the image. Contour description which is local by the nature of the data can guide segmentation. But this has to be distinguished from the case when the information that could give rise to low level models is not present. A good example are the well known phenomena of illusory contours in human perception. We can perceive solid shapes although a large part of boundary lines physically do not exist. Though perceptual shape resulting from subjective contours or illusions is not our concern in this research since we are dealing with physical shape only, the observation is relevant. In conflicting situations information has to be reorganized and the control system adapted. Also, in simple situations like that in figure 3 contours may not give exact site for segmentation. True, the pair of concavities in the contour segment the contour into two parts belonging to two distinct parts in
3-D, they do not provide a mechanism to segment the 3-D object as such. Indeed, partitioning into relevant parts requires surface boundaries (figure 3, shown in the mean curvature sign map). This example presents the case for not relying entirely on contour information for 3-D segmentation, although contour level segmentation from the same information is correct. Also, discontinuities in surfaces may not project as discontinuities in the planar contour. Thus, the control module has to account for disagreement among primitives, by choosing the one that is most plausible under single viewpoint.

A pertinent issue to address at this time is are we doing too much by simultaneously describing shape at three levels? Is there some way of recognizing the dimensionality of the scene and applying only the primitives needed to the scene? It is true that in a restricted domain, dimensionality is known and an elaborate approach is not needed. We are proposing a general approach that is not tied to a domain of particular dimension. It is certainly possible to recognize some aspects of shape by low-level models, and adapt the control structure accordingly. If all the objects are in the scene are flat, then description can be achieved in terms of only contour primitives, though flat models exist in superquadric vocabulary. Surface models are not at all needed. But the superquadric models will still provide a global region-based shape measure that is not possible to obtain with our contour primitive. A typical way of achieving this in our design is to apply all three primitives as usual. The fact that the scene is two-dimensional will be apparent from the results of all the three modules. The control module can then decide not to go for surface segmentation at global level. Let us consider another scenario. If the object has a hole (visible as an occluding contour, figure 6), there is a good probability of not obtaining a superquadric model for it. However, this is not always true, take for example, a box with a cylindrical hole through it. A model for the box exists and is recoverable.

During the segmentation process the control module has also to decide on part/whole (or part/detail) relationships. This requires determining the scale of a potential part given the overall size of the object and deciding to consider it a part or just a detail of the object that can be ignored (implying that current description is adequate). This requires that the global control program must have the resolution of the parameters and thresholds predetermined, or if possible, adjusted during the process. Some of those parameters are the following:

1. The size (or range of sizes) of the local neighborhood for local processing.

2. Acceptable tolerance for error in model evaluation, keeping in view the limitations of shape models.

3. The size and shape of models. When does a circular cylinder become elliptical, or at what angle two planes must meet for a roof edge to exist?
4. The number (or range) of expected segmented units,

5. The thresholds for partitioning and aggregation.

6. The level of details that we wish to explain.

5 Robust Fixed Size Confidence Procedures

5.1 Overview

In order to use robots to (i) explore unknown environments and (ii) assemble or disassemble structures or components of a space station or remote sensing platform, it is critical that these robotic systems be able to effectively integrate or fuse information from multiple sensors and/or information from a mobile sensor. For example, the tasks of assembly and disassembly require that the robots be able to manipulate complex parts which will not be in known initial positions. Thus, the robot must use one or more sensor to locate the parts and estimate their orientations. This, in turn, requires the robots be able to combine data from these multiple position sensors. This is the problem of multi-sensor fusion of positional information.

Our research in active sensing is based on the theory and application of multiple sensors in the exploration of environments which are characterized by significant \textit{a priori} uncertainties. In addition to uncertainty in the environment, the sensors themselves exhibit noisy behavior. While good engineering practice can reduce certain noise components, it is impractical if not impossible to completely eliminate them. Thus, all sensor measurements are to some degree uncertain. However, sensor errors can be modeled statistically, using both physical theory and empirical data; see for example Kanade (1987) and McKendall and Mintz (1986). In developing these models, one recognizes that a single distribution is usually an inadequate description of sensor noise behavior. It is much more realistic and much safer to identify an envelope or class of distributions, one of whose members could reasonably represent the actual statistical behavior of the given sensor. This use of an uncertainty class (or equivalently: an envelope, set, or neighborhood) in distribution space, protects the system designer against the inevitable unpredictable changes which occur in sensor behavior. Reasons for uncertainty in statistical sensor models include: sporadic interference, drift due to aging, temperature variations, miscalibration, quantization, and other significant nonlinearities over the dynamic range of the sensor. See for example Fuma and Bajcsy (1988). The purpose of this research is to examine a sensor fusion problem for both linear and nonlinear location data models using statistical decision theory (SDT). The expected contribution of this research is the application of SDT to obtain: (i) a robust test of the hypothesis that data from different sensors is consistent; and (ii) a robust procedure for combining the data which pass this preliminary consistency test.
Here, robustness refers to the statistical effectiveness of the decision rules when the probability distributions of the observation noise and the \textit{a priori} position information associated with the individual sensors are uncertain. The standard (linear) location data model refers to observations of the form: \( Z = \theta + V \), where \( V \) represents additive sensor noise and \( \theta \) denotes the "sensed" parameter of interest to the observer. The parameter \( \theta \) is referred to as a location parameter, since the distribution of \( Z \) is obtained from the distribution of \( V \) by a translation. While the location parameter fusion problem is only one of many possible fusion paradigms, it does provide a useful starting point for considering more complicated problems, e.g., nonlinear location sensor models of the form: \( Z = h(\theta) + V \), where \( h \) denotes a given (nonlinear) function. It also provides a useful starting point for considering other important generalizations of the location sensor model such as: \( Z = h(\theta + V) \).

The fusion of location data for location estimation has been discussed in the robotics literature by Ayache and Faugeras (1987), Chatila and Laumond (1985), Crowley and Ramparany (1987), Durrant-Whyte (1986, 1988), Matthies and Shafer (1986), and Smith et al. (1986, 1987). In Ayache and Faugeras (1987), Chatila and Laumond (1985), Crowley and Ramparany (1987), and Smith et al. (1986, 1987), the authors assume the sensor noise can either be adequately modeled by Gaussian distributions with known means and covariances, or by distributions characterized only by specified first and second moments. In this latter situation the analysis is limited to affine decision rules which are evaluated on the basis of quadratic loss. Subject to these limitations (affine procedures and quadratic loss), the results of this decision model are "equivalent" to invoking a Gaussian model. Whereas, in Durrant-Whyte (1986, 1988) and Matthies and Shafer (1986), the authors recognize the need to address the existence of non-Gaussian sampling distributions. While Gaussian models offer a degree of mathematical elegance and simplicity, the adoption of Gaussian models for sampling distributions imposes substantial risk on the decision-maker when the actual sampling distributions possess heavy tails, e.g., they exhibit departures from the Gaussian model in the form of \( \epsilon \)-contamination uncertainty classes. This lack of robustness of decision procedures based on Gaussian sampling distributions has been discussed in the statistics literature for more than thirty years. Detailed examinations of the theory and applications of robust statistical inference appear in the monographs by Huber (1981) and Hampel et al. (1986). In addition to heavy-tailed deviations from a Gaussian model, it is useful to consider uncertainty classes defined by less severe departures, e.g., non-Gaussian uncertainty classes which are dominated by given Gaussian distributions.

In the sequel we: (i) delineate several paradigms for robust fusion of multi-sensor linear location data; (ii) introduce some essential nomenclature and definitions from SDT; (iii) review the earlier decision-theoretic results on which this research is based; (iv) delineate a methodology for robust fusion of multi-sensor location data; and (v) describe our current research objectives.
5.2 Paradigms for Sensor Fusion of Location Data

In this subsection we delineate several paradigms for robust fusion of linear location data models. We restrict our attention to observations of one-dimensional location parameters. The results of this one-dimensional analysis can be applied to the multidimensional case by doing a component by component analysis. Alternatively, one can pursue a formal multidimensional extension of the methodology described here.

The general one-dimensional paradigm is delineated as follows. We assume that we are given the sampled outputs of $r$ sensor systems $\{S_i : 1 \leq i \leq r\}$. We denote the $k^{th}$ sampled output of $S_i, 1 \leq k \leq N_i$ by:

$$Z_{ik} = \mu_i + W_i + \theta_i + V_{ik},$$

where:

- $a_i \leq \theta_i \leq b_i$, denotes an unknown location parameter with known bounds $a_i$ and $b_i$.\(^1\) In many applications there is a common interval of location parameter uncertainty for all sensors. However, there is no need to make this assumption in the following mathematical developments.

- $\mu_i$, denotes a known constant (offset) associated with the position of sensor $S_i$ with respect to a common origin.

- $V_{ik}$, denotes the additive observation noise associated with the $k^{th}$ observation (sample) from $S_i$. The random variables $\{V_{ik} : 1 \leq k \leq N_i\}$ are assumed to be independent and identically distributed (i.i.d.). We further assume that the noise process associated with $S_i$ is independent of the noise process associated with $S_j$, when $i \neq j$. Finally, we assume that the probability distribution of $V_{ik}$ belongs to a given uncertainty class of distributions, $\mathcal{F}_i$. We do not assume that the noise processes associated with different sensors are identically distributed.

- $W_i$, denotes the uncertainty in the position of sensor $S_i$ with respect to a common origin. We consider two cases: (i) the position uncertainty of $S_i$ can be expressed by a known interval $[l_i, u_i]$ — with no \textit{a priori} probabilistic description; or (ii) the position uncertainty of $S_i$ can be expressed by an unknown probability distribution from a given uncertainty class $\mathcal{P}_i$. In each case, we assume that the position uncertainty of $S_i$ is independent of the observation noise $\{V_{ik} : 1 \leq k \leq N_i\}$, and independent of the observation noise and position uncertainty of the other sensors.

---

\(^1\)The bounds $a_i$ and $b_i$ may assume infinite values.
Remark 5.1 Without loss of generality, we can assume that the known offsets \( \{\mu_i : 1 \leq i \leq r\} \) are each zero, since nonzero values can be subtracted from the observations \( \{Z_{ik}\} \). Further, if the known, generally asymmetric, interval of uncertainty \([a_i, b_i]\) in \( \theta_i \) is finite, then the observations \( \{Z_{ik}\} \) can be shifted and the interval of uncertainty \([a_i, b_i]\) can be replaced by \([-d_i, d_i]\), where \( d_i = (b_i - a_i)/2 \). Similarly, we can assume the interval of sensor position uncertainty (where applicable) is again symmetric. Thus, (5.1) can be replaced by:

\[
Z_{ik} = W_i + \theta_i + V_{ik},
\]

where: \( |\theta_i| \leq d_i \), and (where applicable) \( |W_i| \leq \eta_i, 1 \leq i \leq r \).

The uncertainty classes \( \mathcal{F}_i \) and (where applicable) \( \mathcal{P}_i, 1 \leq i \leq r \), denote neighborhoods in the space of probability distributions which are deemed to adequately characterize the uncertainty in the specifications of the sampling distributions. Models for several uncertainty classes are described in Subsections 5.4, 5.5, and 5.6.

In terms of this paradigm, our research goals become the delineation of: (i) a robust test of the hypothesis that data from different sensors is consistent, i.e., testing the hypothesis that \( \theta_i = \theta_j, 1 \leq i < j \leq r \); and (ii) a robust procedure for combining the data which pass this preliminary consistency test.

In the following subsection, we introduce the concepts of robust minimax decision rules and robust confidence procedures. These concepts provide the basis for the discussion in the sequel.

5.3 Nomenclature and Definitions from SDT

The standard statement of a minimax location parameter estimation problem includes as given: a parameter space \( \Omega \); a space of actions \( \mathcal{A} \); a loss function \( L \) defined on \( \mathcal{A} \times \Omega \); and a cumulative distribution function (CDF) \( F \); see for example Berger (1985). If the underlying CDF is imprecisely known, then this standard minimax decision model must be reformulated to account for this additional uncertainty. Statistical decision rules which are applicable in this more general problem setting are referred to as robust procedures.

This research report addresses the delineation of robust fixed size confidence procedures for a restricted parameter space. These robust confidence procedures are based, in turn, on the solution of a related robust minimax decision problem:

Let \( Z \) denote a vector of \( N \) i.i.d. observations of a scalar random variable with CDF \( F(z - \theta) \), where \( F \in \mathcal{F}, \) a given uncertainty class. Let \( \Omega = \mathcal{A} = [-d, d] \), and define a zero-one loss function \( L \) on \( \mathcal{A} \times \Omega \):

\[
L(a, \theta) = \begin{cases} 
0, & |a - \theta| \leq \varepsilon; \\
1, & |a - \theta| > \varepsilon;
\end{cases}
\]
where $e > 0$, is given. Further, let $R(\delta, \theta, F) = E[L(\delta, \theta) | \theta, F]$ denote the risk function of the decision rule $\delta$ given $\theta \in \Omega$ and $F \in \mathcal{F}$.

**Definition 5.1** An estimator $\delta^*$ is said to be a robust minimax estimator for $\theta$, if for all $\delta$:

$$\sup_{\delta \in \Omega} R(\delta^*, \theta, F) \leq \sup_{\delta \in \Omega} R(\delta, \theta, F).$$

**Observation 5.1** The connection between the robust minimax rule $\delta^*(Z)$ and a robust fixed size confidence procedure is obtained by noting that:

$$C^*(Z) = [\delta^*(Z) - e, \delta^*(Z) + e] \quad (5.3)$$

can be interpreted as a robust confidence procedure of size $2e$ which has the highest confidence coefficient $\inf_{\theta, F} P_{\theta, F}[\theta \in C^*(Z)]$.

Subsections 5.4, 5.5, and 5.6 of this report are organized as follows: Subsection 5.4 presents the results of previous research on robust fixed size confidence procedures; Subsection 5.5 describes a methodology for robust fusion of multi-sensor (linear) location data based on the theory and application of robust fixed size confidence procedures; and Subsection 5.6 delineates our current research objectives.

**5.4 Previous Research**

The paper by Zeytinoglu and Mintz (1988) addresses the questions of the existence, structure and behavior of robust fixed size confidence procedures defined by (5.3). These results are based on earlier research on optimal fixed size confidence procedures (Zeytinoglu and Mintz, 1984). The main results of this previous research requires Definitions 5.1 - 5.3, and are summarized by Theorems 5.1 and 5.2. For brevity, we restrict our discussion here to the case when $d/e$ is an integer $\geq 2$.

**Definition 5.2** Let $\mathcal{F}$ denote an uncertainty class with upper-envelope CDF $F_u$:

$$\mathcal{F} = \{F: F(x^-) \leq F_u(x), x \leq 0; \text{ and } F(x) \geq F_u(x), x > 0\}, \quad (5.4)$$

where $F_u$ has a density function which is unimodal and symmetric about zero.$^2$

Definition 5.1 makes precise the notion of an uncertainty class $\mathcal{F}$ in the space of cumulative distribution functions (CDF's) which model the statistical behavior of the given sensors. In general, an uncertainty class may contain non-Gaussian, asymmetric, and discontinuous CDF's.

$^2$F(x$^-$) denotes the left-hand limit. The nomenclature upper-envelope refers to the bounding behavior of $F_u$ for $x \leq 0$. 

22
Remark 5.2 We allow $F_u$ to be substochastic, i.e., $F_u$ can have less than unit probability mass. Thus, the usual $\epsilon$-contamination models can be represented by $F$ (5.4).

Definition 5.3 Let $C$ denote the class of nonrandomized, odd, monotone nondecreasing decision rules $\delta: E^1 \to A$. Let $\Delta \subset C$ denote the set of rules $\delta(t)$, defined for $t \geq 0$ by:

$$
\delta(t) = \begin{cases} 
  d - e, & c + a_n + 2ne \leq t; \\
  . & \\
  . & \\
  . & \\
  t - a_2, & c + a_2 + 2e \leq t < c + a_2 + 4e; \\
  2e + c, & c + a_1 + 2e \leq t < c + a_2 + 2e; \\
  t - a_1, & c + a_1 \leq t < c + a_2; \\
  c, & c \leq t < c + a_1; \\
  t, & 0 \leq t < c;
\end{cases}
$$

where: $0 \leq a_1 \leq a_2 \leq \ldots \leq a_n < \infty$, $d = (2n + 1)e + c$, and $c$ equals zero $(e)$ if $d$ is an odd (even) multiple of $e$.

Definition 5.2 identifies the class of decision rules $\Delta \subset C$ on which the robust fixed size confidence procedures (5.3) are based.

Theorem 5.1 (Zeytinoglu and Mintz, 1988) If $N = 1$, and $F$ denotes the uncertainty class (5.4) with upper-envelope $F_u$, then there exists a robust $C$-minimax rule $\delta^* \in \Delta$, if $e \geq B(d/e, N = 1, F_u)$ — a given bound. Further, if $F_u$ possesses a monotone likelihood ratio, then $\delta^*$ is a robust minimax Bayes rule.

Theorem 5.1 addresses the single-sample decision problem. Theorem 5.2 extends the robust $C$-minimax results of Theorem 5.1 to the multi-sample problem by restricting the class of estimators to rules of the form $\delta(T(Z))$, where: $\delta \in C$, $T: E^N \to E^1$, and $T(Z)$ possesses a density function which is unimodal and symmetric about $\theta$. Examples of candidate $T$ statistics include: the sample mean, the sample median, and other symmetric linear combinations of order statistics. For brevity, we restrict our discussion here to the sample median.

Definition 5.4 Let $Z_M$ denote the median of the $N$ observations $Z$. The decision rule $\delta^*(Z_M)$, defined by the composition $\delta^* \circ Z_M$, is said to be a median-minimax estimator for $\theta$, if $\delta^*$ is a minimax rule in the usual sense. The respective definitions of robust median-minimax rules, and robust $C$-median-minimax rules are obtained as before.

---

3 Due to the existing symmetry, all function definitions are stated for nonnegative arguments.
4 A rule is (robust) $D$-minimax if it is (robust) minimax within the class $D$.
5 If $N$ is even, $Z_M = (Z_{[N/2]} + Z_{[N/2] + 1})/2$. 

23
Theorem 5.2 (Zeytinoglu and Mintz, 1988) If $N > 1$, and $\mathcal{F}$ denotes the uncertainty class (5.4) with upper-envelope $F_u$, then there exists a robust $C$-median-minimax rule $\delta^* \in \Delta$ if $e \geq B(d/e, N, F_u)$ — a given bound. Further, if the upper-envelope CDF of $(Z_M - \theta)$ possesses a monotone likelihood ratio, then $\delta^*$ is a robust median-minimax median-Bayes rule.

We conclude this subsection with an example based on Theorem 5.1.

Example 5.1 (an $\epsilon$-contamination model) Let $d = 3e$, and $\mathcal{F}$ denote the uncertainty class:

$$\mathcal{F} = \{F : F = (1 - \epsilon)\Phi + \epsilon H\},$$

where: $\Phi = \mathcal{N}(0, 1)$, the CDF $H$ is symmetric about zero, and $0 < \epsilon < 1/2$. The corresponding (substochastic) upper-envelope is:

$$F_u = (1 - \epsilon)\Phi + \epsilon/2.$$

In this example $B(d/e, N = 1, F_u)$ is:

$$B(d/e, N = 1, F_u) = -(1/2)F_u^{-1}(1/4) = -(1/2)\Phi^{-1}((1 - 2\epsilon)/(4 - 4\epsilon)).$$

The $C$-minimax rule $\delta^*(Z)$ is:

$$\delta^*(Z) = \begin{cases} 
2e, & a_1 + 2e \leq Z; \\
Z - a_1, & a_1 \leq Z < a_1 + 2e; \\
0, & 0 \leq Z < a_1;
\end{cases} \tag{5.5}$$

where $a_1$ satisfies:

$$F_u(a_1 - \epsilon) = 2F_u(-a_1 - \epsilon),$$

or equivalently,

$$\Phi(a_1 - \epsilon) = 2\Phi(-a_1 - \epsilon) + \epsilon/(2 - 2\epsilon).$$

Thus, if $e \geq B$, then $\delta^*(5.5)$ is a robust $C$-minimax rule for this $\epsilon$-contamination model.

This solution is easily extended to other values of $d/e$ and nominal distributions. The required calculations include: the computation of the vector $a$ which parameterizes the underlying $C$-minimax rule $\delta^*$, and the computation of the bound $B(d/e, N = 1, F_u)$ — which are each readily obtained by means of a Newton-Raphson algorithm.
5.5 A Methodology for Robust Multi-Sensor Fusion

5.5.1 Preliminary Remarks

In this subsection we describe a methodology for robust fusion of multi-sensor (linear) location data based on the theory and application of robust fixed size confidence procedures. Our approach contains two distinct phases:

- **Phase I** provides a test of the hypothesis that the location data (5.2) from sensor $S_i$ is consistent with the location data from sensor $S_j$, where $i < j$, i.e., we test the hypothesis that $\theta_i = \theta_j$, $i < j$.

- **Phase II** provides a means of combining the location data from the individual data sets which "pass" the Phase I test, i.e., those deemed to be consistent.

In both phases of this process, we seek procedures which are robust to heavy-tailed deviations from the nominal sampling distribution, such as exhibited in $\epsilon$-contamination uncertainty classes. Our usage of the terminology "robust" is also intended to imply that the procedures have satisfactory behavior when the actual sampling distribution coincides with the nominal, e.g., a given Gaussian distribution.

5.5.2 Sample Sizes and Uncertainty Classes

In developing suitable consistency tests, there are three classes of sample sizes to address: (i) the single sample case, $N_i = 1$; (ii) the small sample case, $1 < N_i \leq 20$; and (iii) the large sample case, $N_i > 20$. In defining these classes, it is important to observe that: (i) The transition ($N = 20$) between the small sample and large sample cases is not a precise threshold value — the appropriate selection of this threshold is dependent on the uncertainty classes which define the given decision problem; and (ii) The sample sizes $N_i$ and $N_j$ can belong to different sample size domains.

The selection of appropriate sensor noise uncertainty classes $\{\mathcal{F}_i : 1 \leq i \leq r\}$ is an important issue in the development of a methodology for robust fusion of multi-sensor location information. Since, at the minimum, we seek to account for the occurrence of noise distributions with heavy tails, it is appropriate to consider $\epsilon$-contamination uncertainty classes. In the sequel, we adopt a unimodal symmetric $\epsilon$-contamination model $\mathcal{F}_{\epsilon_i}$ for each sensor $S_i$, $1 \leq i \leq r$. In particular, we adopt the $\epsilon_i$-contaminated Gaussian model for sensor $S_i$ which is defined by:

$$\mathcal{F}_{\epsilon_i} = \{F : F = (1 - \epsilon_i)\Phi + \epsilon_i H\},$$

where: $\Phi = \mathcal{N}(0,1)$, the CDF $H$ is unimodal and symmetric about zero, and $0 < \epsilon_i < 1/2$.  

25
It is also necessary to model the \textit{a priori} position uncertainty in each sensor. For brevity, we restrict our discussion here to non-stochastic models of the form: $|W_i| \leq \eta_i$, $1 \leq i \leq r$, where $\eta_i \geq 0$ is given.

5.5.3 Phase I — Robust Consistency Tests

The following procedure provides a robust test of the hypothesis that $\theta_i = \theta_j$, $i < j$.

\textbf{Case 1:} ($\eta_i = 0$) Let $\mathcal{M}_i$, $1 \leq i \leq r$, denote the class of CDF's defined by the centered sample median $Z_{M_i}$ of $N_i$ i.i.d. samples with CDF $F \in \mathcal{F}_i$. Let $\mathcal{M}_{ij}$, $1 \leq i < j \leq r$, denote the class of CDF's defined by the difference of the centered sample medians $(Z_{M_i} - \theta_i) - (Z_{M_j} - \theta_j)$, where the CDF's of the centered sample medians $(Z_{M_i} - \theta_i)$ and $(Z_{M_j} - \theta_j)$ belong, respectively, to $\mathcal{M}_i$ and $\mathcal{M}_j$. It follows from these definitions that the class $\mathcal{M}_{ij}$ is a set of symmetric unimodal distributions. Further,

$$Z_{M_i} - Z_{M_j} = \theta_i - \theta_j + \nu_{ij},$$

where: the CDF of $\nu_{ij}$ belongs to $\mathcal{M}_{ij}$; and the \textit{a priori} uncertainty in $\theta_i - \theta_j$ is given by the interval $[-d_{ij}, d_{ij}]$, where $d_{ij} = d_i + d_j$.

Hence, we can construct a robust fixed size (2$\epsilon$) confidence procedure for $\theta_i - \theta_j$. The parameter $\epsilon$ is selected by the decision maker and denotes his tolerance to small errors between $\theta_i$ and $\theta_j$. The desired procedure $[\delta^* - \epsilon, \delta^* + \epsilon]$ is based on Theorems 5.1 and 5.2. Finally, the test of the hypothesis $\theta_i = \theta_j$ is obtained as follows: we reject $\theta_i = \theta_j$ if $0 \not\in [\delta^* - \epsilon, \delta^* + \epsilon]$. From this test we also obtain the probability that $\theta_i - \theta_j \in [\delta^* - \epsilon, \delta^* + \epsilon]$. Examples of applications of this class of robust consistency tests appears in McKendall and Mintz (1988).

\textbf{Case 2:} ($\eta_i > 0$) If the uncertainties $\eta_i$, $1 \leq i \leq r$, are suitably small, then the previous test can be applied with only a small modification in the definition of the uncertainty class $\mathcal{M}_{ij}$. The details appear in McKendall and Mintz (1988).

5.5.4 Phase II — Robust Fusion of Consistent Multi-Sensor Location Information

The following procedure provides a robust estimate of the common location parameter $\theta$ of $r$ sensor data sets, $r \geq 3$. We observe at the outset that, when $V_1$ and $V_2$ possess very heavy tails, in general, it is not useful to combine two observations of the form:

$$Z_1 = \theta + V_1$$

$$Z_2 = \theta + V_2$$

by convex combination. For example, if $V_1$ and $V_2$ are independent Cauchy $C(0,1)$ random variables, then any convex combination of $Z_1$ and $Z_2$ will be a $C(\theta,1)$ random variable. Further,
there are random variables with continuous unimodal symmetric density functions whose sample mean has greater variability than any of its $N$ i.i.d. components, for any sample size $N > 1$. (See for example, Brown and Tukey (1946).)

**Case 1:** $(\eta_i = 0)$ Let $\{Z_{M_i} : 1 \leq i \leq r\}$ denote the sample medians of $r$ consistent data sets with common location parameter $\theta$. In order to simplify the exposition, we further assume that the $r$ sample medians are identically distributed. Let $Z_{M_A}$ denote the median of the $\{Z_{M_i} : 1 \leq i \leq r\}$. Let $\mathcal{M}_A$ denote the uncertainty class of the centered sample median $Z_{M_A} - \theta$. Each CDF $F \in \mathcal{M}_A$ is unimodal and symmetric about zero. Thus, we can apply Theorem 5.1 to obtain a robust fixed size confidence procedure $[\delta^* - \epsilon, \delta^* + \epsilon]$ for $\theta$. Examples of applications of this class of confidence procedures for the robust fusion of consistent multi-sensor location information appears in McKendall and Mintz (1988).

**Case 2:** $(\eta_i > 0)$ If the uncertainties $\eta_i$, $1 \leq i \leq r$, are suitably small, then the previous robust confidence procedure for estimating $\theta$ can be applied with only a small modification in the definition of the uncertainty class $\mathcal{M}_A$. The details appear in McKendall and Mintz (1988).

### 5.6 Current Research Objectives

Our research efforts in robust multi-sensor fusion span: (i) the experimental application of our current algorithmic developments in robust multi-sensor fusion and (ii) the extension of this current theory. This research is an integral part of the machine perceptual development – active sensing paradigm, where robotic systems are equipped with multiple sensors.

**Experimental Studies in Robust Multi-Sensor Fusion** We are studying the application of the current concepts and techniques described in Subsection 5.5 to the domain of robust multi-sensor fusion. This research includes the empirical determination of uncertainty classes which are defined by upper-envelope CDF's which are not necessarily unimodal nor symmetric. These experimental studies are based on position data from vision, acoustic and tactile sensors which are currently available in the GRASP laboratory. Further, as new theoretical developments (described below) become available, these will be incorporated in related experimental studies.

**Randomized Robust Confidence Procedures** One of the sufficiency conditions in the theory of robust fixed size confidence procedures delineated in Subsection 5.4 of this report is that the value of $\epsilon \geq B$ — a given bound. The delineation of a complete theory requires that we also consider the case where $\epsilon < B$. It has been shown, by Martin and Mintz (1985), that where $\epsilon < B$, it is necessary to consider randomized decision rules to obtain robust fixed size confidence procedures. We are developing a complete theory of robust fixed size confidence procedures by including randomized decision rules.
Finite Decision Models  It is important to consider finite decision models, i.e., those decision models where $\Omega$ is a finite set. Such models arise in the study of nonlinear sensors which are subject to quantization. We are developing a corresponding theory for these finite decision models. In developing this theory we are also considering noise distributions which are not absolutely continuous. Preliminary results in this direction (McKendall and Mintz, 1988) demonstrate that it is necessary to consider nonmonotone decision rules to obtain a complete solution.

Nonlinear Location Models  Because of quantization, saturation and other nonlinear behavior, many sensors cannot be adequately modeled by linear $(Z = \theta + V)$ location models. We are developing a theory of robust confidence procedures to account for nonlinear location models of the form: $Z = h(\theta) + V$, where $h$ is a given nonlinear function. We are also considering this same problem when $h$ is uncertain, i.e, $h \in \mathcal{H}$ — a given uncertainty class.

Other Nonlinear Models  Because of the existence of output nonlinearities, many sensors cannot be adequately modeled by nonlinear location models of the form: $Z = h(\theta) + V$. We are developing a theory of robust confidence procedures to account for sensor output nonlinearities of the form: $Z = h(\theta + V)$, where $h$ is a given nonlinear function. We are also considering this same problem when $h$ is uncertain, i.e, $h \in \mathcal{H}$ — a given uncertainty class.

Multivariate Models  We are considering alternative approaches to the multivariate problem which address the problem directly, as opposed to a component by component analysis.

6 References


