August 1990

Image Understanding at the GRASP Laboratory

Ruzena Bajcsy

University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/cis_reports

Recommended Citation
Ruzena Bajcsy, "Image Understanding at the GRASP Laboratory", August 1990.


This paper is posted at ScholarlyCommons. http://repository.upenn.edu/cis_reports/577
For more information, please contact libraryrepository@pobox.upenn.edu.
Image Understanding at the GRASP Laboratory

Abstract
Research in the GRASP Laboratory has two main themes, parameterized multi-dimensional segmentation and robust decision making under uncertainty. The multi-dimensional approach interweaves segmentation with representation. The data is explained as a best fit in view of parametric primitives. These primitives are based on physical and geometric properties of objects and are limited in number. We use primitives at the volumetric level, the surface level, and the occluding contour level, and combine the results. The robust decision making allows us to combine data from multiple sensors. Sensor measurements have bounds based on the physical limitations of the sensors. We use this information without making a priori assumptions of distributions within the intervals or a priori assumptions of the probability of a given result.

Comments
Image Understanding
at the
GRASP Laboratory

MS-CIS-90-52
GRASP LAB 225

Ruzena Bajcsy

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

August 1990

ACKNOWLEDGEMENTS:
Navy Grant N0014-88-K-0630, AFOSR Grants 88-0244, AFOSR
88-0296; Army/DAAL 0389-C-0031PRI; NSF Grants CISE/CDA
88-22719, IRI 89-06770; and DuPont Corporation.
Abstract

Research in the GRASP Laboratory has two main themes, parameterized multi-dimensional segmentation and robust decision making under uncertainty. The multi-dimensional approach interweaves segmentation with representation. The data is explained as a best fit in view of parametric primitives. These primitives are based on physical and geometric properties of objects and are limited in number. We use primitives at the volumetric level, the surface level, and the occluding contour level, and combine the results. The robust decision making allows us to combine data from multiple sensors. Sensor measurements have bounds based on the physical limitations of the sensors. We use this information without making a priori assumptions of distributions within the intervals or a priori assumptions of the probability of a given result.

1 Introduction

Our basic approach to Image Understanding can be summarized as follows: we seek to divide the observed scenes into 3D objects. 2D images are observations/measurements of these 3D physical objects under certain illuminations and perspective projections. Hence, the process of Image Understanding includes the transformation of the 2D data into a description of 3D objects in terms of physical and geometric primitives. This grouping of image data, to produce such a description, is called segmentation.

Though image segmentation has been treated separately from shape representation in the past, solving the two problems separately is very difficult. If an image is correctly divided into 3D components, ambiguities in the 2D segmentation can be resolved more easily. Conversely, a range image can be partitioned more easily into 3D shapes given a good 2D segmentation of the same scene. Indeed, dividing the scene into 3D primitives is a form of segmentation. Hence, 2D image segmentation and 3D description of parts should act together as a cooperative process [Bajcsy et al., 1990], a combined 2D–3D segmentation process.

Range images and 2D images complement each other. Range data is a reflection of geometric properties of the objects only. However, 2D images contain indirect information about geometric properties in combination with surface properties (color, texture, translucence, etc.), which is not available at all in the range data. Because of this, a cooperative 2D/3D segmentation process has the potential to be a major improvement over separate segmentation.

By extension, the grouping of data through time is also segmentation. One aspect of grouping temporal data is the construction of range images via sliding stereo or structured light range images. Another aspect of grouping temporal data is the understanding of movable or removable parts in objects in a scene. Understanding that a part moves as a unit, separately from another part, is a partitioning of the data. Without a priori knowledge, this cannot be detected by noncontact sensing. This segmentation would allow understanding of the dynamic properties, the mechanical properties and the kinematic properties of objects. However, in this paper, we concentrate on the aspects of 2D/3D segmentation process which can be accomplished through non-contact sensing.

1. Definition of the segmentation problem:

- Segmentation is partitioning the space into meaningful parts. This can be done on intensity images or range images.
- Segmentation is data reduction and requantization of sensory measurements into some primitive elements.
- Segmentation is an inference of a symbolic description of the world from one or more images.

2. What kind of measurements are available in noncontact sensing?

- Multispectral Image: camera with filters having different spectral sensitivities. We measure energy flux incident on one image plane (irradiance) that combines the following components: energy and spectral distribution of the incident light,
the following discontinuities: depth, orientation, albedo, shadow, shading, and specularity. Hence, our segmentation and resulting descriptions will be in terms of physical properties of the world (surface reflectance, shading, shadow, highlights and geometry) rather than in terms of image attributes.

3. What kind of primitive elements should we use?
- Primitives must balance the trade-offs between:
  - data reduction versus faithfulness to measured data and
  - localness versus globalness.
- Primitives should correspond to segments in terms of physical phenomena.

Physical phenomena are manifested in the image via the following discontinuities: depth, orientation, albedo, shadow, shading, and specularity. Hence, our segmentation and resulting descriptions will be in terms of physical properties of the world (surface reflectance, shading, shadow, highlights and geometry) rather than in terms of image attributes.

In Section 2 our work detecting highlights and inter-reflections is described. Section 3 details segmentation of images into objects/surfaces made of specular and diffuse materials [Bajcsy et al., 1989]. Section 4 explains the segmentation of the range image and/or lightness image into surfaces. Section 4 represents our efforts to recover underlying geometric structure from an image. Since geometry involves more than just surfaces [Leonardis et al., 1990], we find volumetric descriptions in terms of superquadric parts. All of this processing is only from one viewing angle, which is not sufficient for describing a scene composed of opaque objects. This question of where to go next, based on the first view, is described briefly in Section 5 and in more detail in the paper by Maver and Bajcsy in this Proceeding.

The above work is being extended into the development of a formal model for an observer of an indoor scene being explored by a mechanical hand [Bajcsy and Sobh, 1990]. The task for the observer is to have a full view of the hand and the object being held and/or manipulated. We have adopted the formalism of discrete event dynamic systems (DEDS), which allows us to predict under which conditions the observer-automaton can accomplish the task. This includes both the stability and the observability conditions. The formalism of DEDS has been described in [Ho, 1987], [Ramadge and Wonham, 1987b; Ramadge and Wonham, 1987a] and others.

Observing and understanding motion is an established part of the Image Understanding effort. In the GRASP Laboratory, we have used a temporally-oriented approach rather than a spatially oriented approach. This approach is described in Section 6, and detailed in the paper by Wohn and Iu in this Proceeding.

The underlying theory of updating procedures and decision making under uncertainty is summarized in section 7, and described more thoroughly in the paper by Mintz, McKendall and Kamberova in this Proceeding.

2 Color Image Understanding: Image Segmentation and Detection of Highlights and Inter-Reflections Using Color.

Color image segmentation should be based on changes in object colors. In addition to the object color changes, however, an image of three-dimensional real objects contains variations because of shading, shadows, highlights and inter-reflections. Detection and separation of highlights for successful segmentation has been the focus of recent efforts in color image segmentation [Gershon, 1987] [Klinker et al., 1988]. We have approached the construction of a computational model for color image segmentation not only with detection of highlights, but also with detection of small color changes induced by inter-reflections.

We use the dichromatic model [Shafer, 1985] for dielectric materials. There are two reflection mechanisms—surface reflection and body reflection. The surface or interface reflection occurs at the interface of air and object material, and can be specular and/or diffuse reflections depending on the surface roughness. Surface reflectance at dielectric materials are spectrally flat for both specular and diffuse reflections. Body reflection, on the other hand, is spectrally colored depending on the pigments, and always diffused. When the body reflectance is non-flat, we can detect the flat component of surface (or interface) reflectance regardless of surface roughness. Highlights are caused by specular surface reflection; inter-reflections between the objects are caused by both surface (specular or diffuse) and body reflections.

To better represent and process the image color, a color metric space is developed based on the physical model of the camera and filters. The measured color in R,G,B space is transformed into the metric space using a set of orthogonal basis functions. We used the first three of Fourier basis functions. Within our framework, however, any orthogonal basis functions can be used for better representation of natural colors [Cohen, 1964] [Judd et al., 1964]. The metric space is similar to the opponent space in human vision with intensity, hue and saturation. With orthogonal values, we can manipulate each component of color separately or in combinations. Transformation from R,G,B into orthogonal values is a pixel parallel process, and is implemented on a Connection Machine (CM2a).

Since illumination is usually spectrally colored, calibration of measured images is performed with a white object of reference to whiten the illumination. Whereas the spectral distribution of object surfaces is not changed by shading and shadow, it is affected by highlights and inter-reflections between the objects even under the white illumination. Since the highlights add the whiteness to the object color under the whitened illumination, they can be detected by observing the change of saturation in the uniformly colored objects, which is equiva-
lent to examining to detect any spectrally flat reflectance added in the body color. The inter-reflections are also detected with the change in saturation and in hue values.

Segmentation is carried out in two steps; hue and saturation segmentation under the assumption that objects are piecewise uniform in hue and saturation. Intensity is not easy to use in the presence of shading, shadow and highlights. Illumination whitening is important for hue segmentation since under white illumination, the highlights do not shift the hue values of object colors. Highlights are detected by observation of saturation values. The use of the reference plate is not necessary when the illumination is weakly colored. Roughly detected highlights can be used for the reference. Within each region, the detection and separation of highlights and inter-reflections is highly parallel, since it is accomplished via thresholding and arithmetic operations.

Our method is improved over previous works [Gershon, 1987] [Klinker et al., 1988], in a few significant ways. The previous methods can detect strong surface specular reflection, but are not reliable in detecting small surface reflections that are diffused. Since we interpret the reflection mechanisms in our color space with all the spectral characteristics of sensors considered, we can better observe the spectral variation of reflection. Therefore, we can detect not only strong and distinctively appearing specular reflection, but also small surface reflections which are spatially diffused. Inter-reflections usually have a diffused appearance. The inter-reflections can be detected by the change in saturation values although hue values also change.

Figure 1 (b) shows Body Reflections and (c) shows the Surface Reflections of which the original intensity is shown in Figure 1 (a). Though inter-reflections are barely visible in the original color image, they are not visible in Figure 1 (a). These inter-reflections are visible in figure 1c, as spatially diffused, low-intensity values on the horizontal strip.

In our previous work, we have concentrated on spectral information. We are expanding our investigation to include:

1. Multi-dimensional segmentation. The spatial segmentation results can be used with segmentation by hue and saturation. [Leonardis et al., 1990].

2. Active Camera Movements. This will allow us to observe different moving patterns of surface and body reflections as the observer moves.

3. Changing illumination (when possible) to disambiguate the image variations due to the change of shading/shadow and albedo.

3 Segmentation as the Search for the best Description of the Image in terms of Primitives.

A new paradigm for image segmentation has been developed. We segment images into piecewise continuous patches [Leonardis et al., 1990]. Data aggregation is performed via model recovery in terms of variable-order bivariate polynomials using iterative regression. All the recovered models are potential candidates for the final description of the data. Selection of the models is achieved through a maximization of quadratic Boolean problem. The procedure can be adapted to prefer certain kinds of descriptions (one which describes more data points, or has smaller error, or has lower order model). We have developed a fast optimization procedure for model selection. The major novelty of the approach is in combining model extraction and model selection in a dynamic way. Partial recovery of the models is followed by the optimization (selection) procedure where only the “best” models are allowed to develop further. The results obtained in this way are comparable with the results obtained when using the selection module only after all the models are fully recovered, while the computational complexity is significantly reduced. We test the procedure on real range and intensity images.

We believe this segmentation schema is a tool that will prove useful in many tasks of early vision. The two procedures (model recovery and recover-and-select) clearly show that the whole can be greater than the sum of its parts (synergism). The iterative approach combin-
ing data classification and model fitting shows that seg-
mentation and modeling are not two independent pro-
cedures but have to be integrated. The procedure which
dynamically combines model recovery with model selec-
tion proves to be much more efficient than applying the
modules one after another.

Another important conclusion that we have drawn
from our work is that reliable segmentation can only be
achieved by considering many competitive solutions and
choosing those which reveal some kind of structure in
terms of underlying models. Fine-tuning of feature de-
tectors does not lead to reliable segmentation, because of
the variability of the input data. Initial local estimates,
no matter how good they are, do not necessarily lead to
a good result, and more global information is needed.
Optimization performed on the level of primitives rather
than on a pixel level not only improves the performance
enormously in terms of computational complexity but
also gives more reliable results.

The results are grouped such that the top row of the
figure (from left to right) shows the original image, its
3-D perspective plot, the reconstructed image from the
piecewise continuous segmented patches, and the 3-D
plot of the reconstructed image. The range images are
displayed with the depth value at each pixel from a re-
ference plane appearing larger if the pixel is closer to the
camera. The white square in the patch indicates the seed
region for that patch. The individual surface patches are
displayed in the second row of the figures in the order
in which they were selected by the model selection pro-
cedure, and are referred to below with their position in
the row, counting from left to right.

The Coffee-mug image: The convex and concave por-
tions of the body of the cup are recovered as individual
second-order patches, as shown in the first two images
of the bottom row in figure 2. The handle consists of
very curved patches which are modeled piecewise for the
given scale (which directly relates to the compatibility
constraint). According to the results, the missing parts
are better described as individual pixels than as para-
metric patches (due to the scale consideration). It should
be noted that the jump (C1) discontinuities are clearly
delineated by the neighboring regions.

4 Integrated approach to 3D shape
(volume, surface, contour) recovery
via parametric descriptions

In section 3, we described segmentation of a 2-1/2 D
image into surface patches. In other work [Solina and
Bajcsy, 1990] we have used parametric descriptions of
superquadrics to fit volumetric aspect of a shape. We
are now in the process of developing a paradigm for de-
composition of complex objects in range images into the
constituent parts based on the shape, using contour, sur-
face, and volumetric primitives [Gupta and Bajcsy, 1990]

Unlike previous approaches, we use geometric prop-
erties derived from both boundary-based (surface contours
and occluding contours), and primitive-based (biquadric
patches and superquadric models) representations to de-
fine and recover part-whole relationships, without a pri-
or knowledge about the objects or the object domain.
The descriptions thus obtained are independent of posi-
tion, orientation, scale, domain and domain properties,
and are based purely on geometric considerations. Since
both boundary-based and primitive-based primitives are
included in our vocabulary, the representation is expres-
se and robust.

In the computer vision literature, the partitioning of
images and description of individual parts is called seg-
mentation and shape representation respectively. We
have presented arguments in Bajcsy, Solina, and Gupta
[Bajcsy et al., 1990] that the problems of segmentation
and representation are related and must be treated simul-
taneously. We propose that for obtaining a global shape
description from single-viewpoint 3-D data requires ad-
dressing shape at the following levels:

1. Volumetric level: Superquadric shape primitives
capable of modeling parts in three dimensions are
needed to describe global 3-D shape of parts.
2. Surface level: Surface primitives describe internal
surface boundaries and surface patches which are
difficult to model by volumetric primitives, but
present s more accurate and detailed description of
shape that is neither too local nor too global.
3. Occluding Contour level: The Occluding con-
tour encodes the 3-D shape of parts projected on
the image plane.

Given the three different modules for extracting vol-
ume, surface and boundary properties, how should they
be invoked, evaluated and integrated? To incorporate
the best of the coarse to fine and fine to coarse segmen-
tation strategy, we perform volume, surface, and bound-
ary fitting in parallel on the input data. This requires
evaluation and comparison of information embedded in
models built by different aggregation methods. The oc-
cluding contour is segmented into parts at concavities
and convexities using the classical techniques. The sur-
face is segmented into planar and bi-quadratic patches us-
ing the segmentation algorithm outlined in the previ-
ous section. The segmentation also gives reliable internal
C1 (orientation) discontinuities, which are vital for
part-segmentation, but are very difficult to localize using
standard edge-detection techniques. The superquadric
model, being an object centered global part-model, is
not amenable to such segmentation techniques. So the
problem of fitting or recovering the superquadric mod-
els to parts of an object has to be attempted in such a
way as to make use of the segmentation information from
the surface and contour models that provide local seg-
mentation at their respective levels. The superquadric
model recovery itself proceeds from coarse to fine (global
to local), generating residuals and hypotheses about vol-
umetric parts as described below. We are developing a
control module to accomplish this non-trivial task in a
systematic manner.

To satisfy the practical constraints of computability
and robustness, we pose the problem of integration in
terms of evaluation of the intermediate descriptions and
segmentation of the objects in a closed loop process. To
evaluate the superquadric models, we have developed a set of quantitative and qualitative measures, that generate global and local residuals of the models [Gupta et al., 1989]. The qualitative residuals are the regions of underestimation, and surface and contour overestimation. The fundamental principle behind our approach is that these residuals are generated because of the presence of parts (or negative volume), otherwise a correct volumetric model would be obtained during the initial global fit. The regions of contour overestimation guide the localization of the concavities in the occluding contour. The concavities in the occluding contour are in turn used to constrain the superquadric model to fit only a part and not the complete object [Gupta and Bajcsy, 1990]. It essentially provides topological constraints to restrict the model to a part of the object. Although the surface segmentation is complete at the surface level, it does not impose any topological constraints needed for superquadric level of part segmentation. In addition, a mechanism to combine individual surface patches to form the volumetric parts is needed to generate strong hypotheses about potential superquadric models. We have observed that the information from the regions of underestimation (primarily caused due to the parts 'sticking out') makes natural clusters of surface patches forming volumetric parts. Using this simple observation, we have obtained encouraging results on various complex objects.

5 Searching for Additional Information

The task of constructing a volumetric description of a scene from a single image is an underdetermined problem, whether it is a range image or an intensity image. Once the first image has been taken, we develop a strategy to get the additional information that will allow us to complete the volumetric description.

Range images are 2 1/2 D images, where the value of a pixel corresponds to the distance from the observer to the closest point in the scene. Some parts of the scene may be partially occluded by objects. In structured light range images, a pixel with no data corresponds to an occluded area.

At the GRASP Laboratory, we have two structured light range imaging systems. One uses a fixed laser (providing the plane of light), while the objects move on a linear stage. Though this system provides high quality range data, it is limited to a single view of the scene. However, our second system uses a laser-camera pair, mounted on a Puma 560 robot arm, which moves over the scene [Tsikos, 1989]. A range image is acquired by moving the laser-camera pair along a straight line in space, aiming at the scene. It is capable of viewing a scene within the robot workspace from many different angles. This capability allow us to construct a complete 3D model of a scene.

Our strategy is to use the information in a narrow zone around the occluded regions. Occluded regions are approximated by polygons. From the height of the border of the occluded regions and from the geometry of the edges of the polygonal approximation, we calculate the direction which will most effectively show us what is in the occluded area. Experimental results on range data are described in the paper by Maver and Bajcsy in this...
6 Image Motion Analysis: A Temporally-oriented Approach

The fact that the relative motion between the viewer and the object can be recovered from a sequence of images is well-known, and articles on the subject are abundant. The majority of previous work dealt with the existence and the uniqueness of solution when the input was assumed to be given in the proper form (image flow field, disparity vectors, position of feature points, conic contours, etc). A typical result states that there are \( K \) solutions for the 3-D motion and the object structure, given \( M \) features over \( N \) frames. For the uniqueness, some suggest that \( K \) could be 1 if \( M' = M \) features are used; others suggest that \( K \) could be 1 if \( N' = N \) frames are used, etc. While these results provide us with some useful theoretical framework, all known algorithms derived from their constructive proofs have turned out to be very sensitive to the input noise [Tsai and Huang, 1984; Waxman and Ullman, 1985].

In estimating 2-D motion, there is no known algorithm that estimates the 2-D image motion with sufficient accuracy for the 3-D motion recovery algorithm. For example, the image flow estimated by any of the well-known techniques [Hildreth, 1984], [Horn and Schunck, 1981] does not seem to be useful at all for the purpose of 3-D motion recovery. Furthermore, the accuracy of image motion estimate is lower-bounded: Even for a noise-free, synthetically generated image sequence, there is the discretization effect in spatial as well as temporal domains. Sometimes this discretization effect alone exceeds the maximal noise level the 3-D motion recovery algorithm can tolerate.

The error in 2-D motion amplifies the error in 3-D motion parameters. The exact nature of error propagation depends not only on the specific algorithm but also on many factors such as the viewing angle of camera, and motion and structure parameters themselves, and thus it is hard to derive the error formula analytically. Unless one imposes unrealistic assumptions such as a huge viewing angle (typical viewing angle of camera lens is 30-50 degrees and the object of interest occupies even a smaller field of view), all the existing algorithms perform poorly under the presence of realistic noise. The difficulty is that the 3-D motion recovery is ill-conditioned.

Hence, a more realistic approach, as far as the recovery of 3-D motion is concerned, is to improve the motion parameters over time, under the presence of noise. Just like any other physical system, the entire process from the image sequence to the 3-D motion may be viewed as a dynamic system, and the problem can be formulated as a non-linear state estimation problem.

Our approach is not merely meant to attempt to improve the motion estimates over time. Most of the previous algorithms mentioned above are “biased” more toward the spatial information in image sequences than the temporal information, in the following sense. In the token matching approach, the main idea is to find the minimal number of points that guarantee a unique solution when a small number (typically 2 or 3) of frames are given. Similarly, the optical flow approach seeks the lowest order of flow derivatives from a single snapshot of optical flow field. Of course, both “instantaneous” approaches have to make use of the temporal information in one way or another, but the extent of temporal information being used is fixed at the stage of problem formulation. We may call these approaches spatially-oriented.

Our approach first explores the temporal information prior to the usage of the spatial one. Here, a typical question one may ask is: “How many frames are needed when \( N \) features are given?” as opposed to “How many features are needed when \( M \) frames are given?” We call it temporally-oriented approach (TOA). The importance of TOA’s is that we can avoid many problems which one may encounter in the SOA’s. Since we observe motion over an extended time interval, we can reduce the number of features that are used in the computation. In fact, we have shown that we could even recover the 3-D motion of a single particle. Consequently, the problem of requiring multiple features can be eliminated and the task of segmentation is thereby reduced. Further, as we use more frames to estimate the 3-D motion, the problem itself becomes more well-conditioned. When we observe a moving object such as a space shuttle or a baseball, the longer we observe, the more accurately we can estimate its motion and predict its position. In TOA’s, we rely on the temporal information from the moving object while keeping the amount of spatial data in a single image as small as possible. Of course we may use multiple features to get a more robust estimate if they are available. Therefore, the TOA’s and the previous multi-frame approach are different in their motivation.

7 Robust Multi-sensor Fusion

We have developed a coherent methodology for fusing data from multiple sensors in uncertain environments. Since sensors exhibit noisy behavior that cannot be eliminated completely, all sensor measurements are uncertain. However, sensor errors can be modeled statistically and geometrically, using both physical theory and empirical data. For example, multiple range sensors may be located on the wings of an aircraft. As the wings themselves flex, the precise location of the sensors move, relative to the center of the aircraft. These sensor location errors can be bounded, but it is not necessarily desirable to characterize these errors statistically. Also, sensors may break or become worn out, but still send data. For example, a camera with a broken IR filter still sends data, which seems to be within the dynamic range of the system (though perhaps at the saturation limit). This kind of uncertainty can be handled by a statistical decision theoretic approach.

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 represent the actual statistical behavior of the given sensor. For example, it may be difficult to fit the error distribution of a
malfunctioning sensor by a Gaussian distribution; however, \( \epsilon \)-contamination models provide a good alternative. Other 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. This use of an uncertainty class in distribution space protects against the inevitable unpredictable changes that occur in sensor behavior. The purpose of this research is to examine sensor fusion problems for both linear and nonlinear location data models using statistical decision theory.

The contributions of this research are the delineation of:

- **Robust tests of consistency of data** from different sensors. This confirms that the measurements are actually measuring the same thing. For example, if two position sensors are actually measuring position of two different objects, then the two resulting data points should not be averaged or combined in any way.

- **Robust procedures for combining data** that pass the preliminary consistency tests.

Robustness refers to the statistical effectiveness of the decision rules when the probability distributions of the observation noise and the precise location of the individual sensors are uncertain. This research provides statistical performance bounds.

While range data is a 1D application of multi-sensor fusion, our current research extends the theories to cover the multi-dimensional case. This allows robust testing of multi-dimensional feature vectors for similarities in non-Gaussian noise, and gives a probability of correctness of the result. This represents a significant theoretical advance in mathematical statistics, since no prior probability distributions are assumed. One example of a multi-dimensional application is color comparison. The color of a pixel can be expressed either as R,G,B (where the data is not independent) or in an orthogonal basis. These theories allow us to compare a pixel against another pixel, and give probability of the colors being the same, without assuming in advance that there is a specific probability of that color occurring.

The decision-theoretic formulation of these problems allows us to find minimax decision rules based on a zero-one loss function. Such rules minimize the maximum probability that the absolute error of estimation is greater than an error tolerance \( \epsilon \). The zero-one loss function is appropriate for situations where a system works if the decision is correct within a given tolerance, and the system fails if the decision is out of bounds.

This research in robust multi-sensor fusion is based on the theory of robust fixed size confidence sets. Our prior work in this area includes the delineation of the existence, structure, and behavior of:

- optimal, nonrandomized, fixed size confidence intervals based on monotone decision rules;

- robust, nonrandomized, fixed size confidence intervals based on monotone decision rules; and

- the extension of these ideas to both randomized, and nonmonotone procedures.


## 8 Conclusion

As it is seen, the Image Understanding Program in the GRASP laboratory represents a coherent programmatic study of material properties, geometric properties, and motion of objects through visual measurements. Our approach is data driven, where we seek the best explanation of the data via parametric models. The result of this approach is not only compact representation but also a measure of goodness of fit which then can be used for feedback correction depending on the task. We also do not depend only on one measurement, but rather try to combine, in a systematic fashion, several different aspects of shape and material. We are in the process of systematically testing our theories with different parameters of illumination, camera positions, signal/noise ratio and complexity of the scene.

## References


