A METHOD FOR INITIAL HYPOTHESIS FORMATION IN IMAGE UNDERSTANDING

Nancy Lehrer
George Reynolds
Joey Griffith

COINS Technical Report 87-04

January 1987

Abstract

This paper presents a method for initial hypothesis formation in image understanding where the knowledge base is automatically constructed given a set of training instances. The hypotheses formed by this system are intended to provide an initial focus-of-attention set of objects from a knowledge-directed, opportunistic image understanding system whose intended goal is the interpretation of outdoor natural scenes. Presented is an automated method for defining world knowledge based on the frequency distributions of a set of training objects and feature measurements. This method takes into consideration the imprecision (inaccurate feature measurements) and incompleteness (possibly too few samples) of the statistical information available from the training set. A computationally efficient approach to the Dempster-Shafer theory of evidence is used for the representation and combination of evidence from disparate sources. We chose the Dempster-Shafer theory in order to take advantage of its rich representation of belief, disbelief, uncertainty and conflict. A brief intuitive discussion of the Dempster-Shafer theory of evidence is contained in Appendix A.

This work has been supported by the following grants: Air Force Office of Scientific Research 86-0021, the Defense Mapping Agency 800-85-C-0012, and the National Science Foundation DCR-8318776.

1. The Interpretation Problem

An important task in image understanding is to develop a mapping from low-level image events (such as regions, lines or surfaces) to higher level semantic abstractions (such as road, grass and foliage). Achieving this task requires developing a knowledge base which defines these semantic abstractions in terms of the low-level image events and using constructive techniques to match or correlate primitive features of the low level events against the knowledge base for each semantic abstraction. An inference technique is required to compare, contrast and combine match scores to create a consistent interpretation. Some schemes for combining information from various sources in initial hypothesis formation include additive "voting" methods [Hans86,Belk86], Dempster-Shafer pooling of evidence [Low82,Wes84,Rey86b], Bayesian methods [Duda73], constraint propagation [Kitc84,Wal72], as well as many ad hoc but heuristically adequate Mycin type systems [Shor76]. In any scheme three questions must be answered: What is the evidence for some inference, How is it represented, and What method will be employed for combining and propagating evidence?

In this paper we present a method for initial object hypothesis formation in image understanding that has evolved from earlier work in the VISIONS research environment documented by Hanson and Riseman in [Hans87]. These hypotheses can then be used as a focus-of-attention set by a knowledge-directed interpretation system and expanded into a more complete interpretation. An automated method is presented for defining a knowledge base based on the frequency distributions of a set of training objects and feature measurements. The system takes an image that is segmented into closed boundary regions and "matches" each region against the stored knowledge base to generate a set of initial hypothesis for a given region. This method can also be used to classify lines, surfaces or any other image abstraction or combination of image abstractions. In our formalism, evidence is represented by a *plausibility function* and combined using a computationally efficient approach to the theory of evidence as pioneered by Glen Shafer referred to as the Dempster-Shafer theory of evidence [Demp67,Shaf76].

At the heart of the Dempster-Shafer formalism is a rich representation of evidence, a belief function or mass function, and a method for combining evidence, Dempster's Rule. The formalism is often criticized for the computational cost associated with Dempster's Rule; in addition the Dempster-Shafer theory does not address the issue of acquiring mass functions. The system presented here addresses both these problems. It is able to use the rich semantics implied by the evidential representation of the Dempster-Shafer formalism without the computational cost associated with Dempster's Rule as well as defining an automatic method for generating a knowledge base. A brief intuitive discussion of the fundamental principles of the Dempster-Shafer formalism are presented in appendix A.

2. The VISIONS Experiments

In general we can think of intermediate-level image abstractions as *tokens* one or two steps removed from the raw image data represented as pixels. Regions can be thought of as area filling abstractions connecting pixels with some homogeneity constraint [Kohl84]. Straight lines may connect and group pixels with the same gradient direction [Burn86,Weis86]. Each intermediate level token can be associated with a feature vector that measures primitive features. If pixels are grouped into closed area-filling regions, measures can be defined which statistically describe a region's color, intensity or texture. Depending upon the line algorithm used, lines also have a variety of primitive features such as length, position in the image (rho), angle or orientation (theta), and measures of the contrast relating the values of pixels on either side of the line.

An approach taken in VISIONS [Hans87] is to use these primitive features to create initial hypothesis labels which associates with each region a semantic label to bootstrap the high-level interpretation process. Although regions are being labeled in this approach, line data can be incorporated into the interpretation process by the use of relationships between lines and regions [Belk86].

2.1 Approaches to Representing and Combining Evidence

Existing approaches to the generation of initial hypotheses have used interactive and heuristic approaches to the knowledge engineering problem. One in particular, the rule-based object hypothesis system of Hanson and Riseman [Hans86], defines constraints on the features of line and region image abstractions; it uses frequency distributions to guide the formation of heuristically defined, piece-wise linear ranking functions called *rules*. In that system, rules provide a vote for a specific object, and a set of "simple" rules are combined via a weighted average to form a "complex" rule which can then be similarly combined. The output of a complex rule represents a weighted combination of evidence from various features and is used to rank order image regions (or collections of regions and lines) according to how well they match a prototype object.

The contribution of the system presented here is two fold: (a) a formal and theoretical foundation for the combination and representation of evidence, and (b) the automatic construction of the knowledge base. It defines the knowledge base automatically using statistical information obtained from a set of training object instances and uses a computationally efficient approach to the Dempster-Shafer theory of evidence for the representation and combination of evidence from disparate sources. The knowledge base is represented in terms of ***plausibility distributions***; their construction takes into consideration the imprecision (i.e. inaccurate feature measurements) and incompleteness (i.e. too few samples) of the statistical information available from the training set. We show that the automatically generated plausibility distributions characterize the range of feature values associated with each object in the training set without biasing the interpretation towards objects more likely to appear at random in the training set. The term plausibility is used to indicate the equivalence between our formalism and the semantics and functionality of the term ***plausibility*** in the Dempster-Shafer formalism.

2.1.1 The Rule System

The task of the rule-based object hypothesis system of Hanson and Riseman [Hans86] system is to provide a set of candidate object hypotheses to a knowledge-directed schema network. Based on these initial hypotheses an island driving focus-of-attention strategy is employed to initiate further processing. Rules are formed by heuristically assigning a vote for a specific object to ranges of feature values. Sets of "simple" rules are combined via a weighted average from "complex" rules. Sets of complex rules are then combined in the same way to form the initial hypotheses. A rule is represented as six threshold values, $\theta_1, \theta_2, ..., \theta_6$, these thresholds define a piecewise linear mapping function from feature space to object space. The intervals $[-\infty, \theta_1]$ and $[\theta_6, \infty]$ represent a veto range, $[\theta_3, \theta_4]$ the range where the object label associated with the prototype feature vector receives a maximum

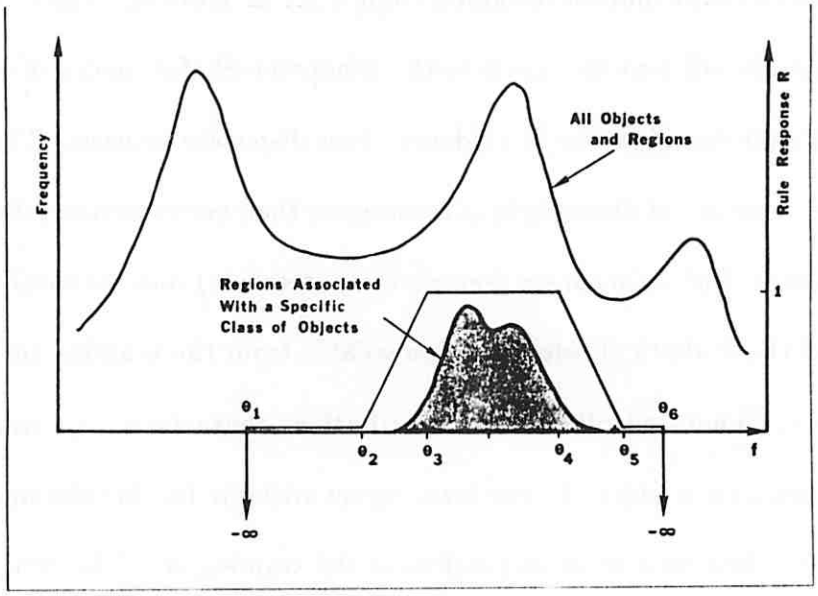

Figure 2.1: Structure of a Simple Rule

Structure of a simple rule for mapping an image feature measurement f into a score for a label hypothesis on the basis of a prototype feature value. The object specific mapping is parameterized by seven values, $f_p, \theta_1, \ldots, \theta_6$.

vote of 1, the intervals $[\theta_1, \theta_2]$ and $[\theta_4, \theta_5]$ a noncommittal vote of zero, and finally $[\theta_2, \theta_3]$ is linearly ramped from 0 to 1 whereas $[\theta_5, \theta_6]$ is ramped from 1 to 0. Figure 2.1 shows the construction of a simple rule. The weights given to the combination rule are heuristically defined on a scale from 1 to 5.

Due to the potentially large number of rules needed for each object in the initial hypothesis set, the rule system is designed to simplify the definition of each rule. The user choosing the threshold values θ_1 through θ_6 for a particular object is equipped with an interactive environment for constructing these rules and displaying their effect. In this system each object has a different set of features which contribute to the objects prototype feature vector. Given an image token, the score for one object is never directly compared against the score for another object; instead for each object, tokens are ranked by how well they score for that object. Regions with the highest scores for a particular object are used as exemplar regions in an island driving strategy applied in later stages of the interpretation.

The rule system was developed in reaction to the problems of using Bayesian techniques for the classification of tokens based strictly on statistical information available from an inadequate training set. In particular, if the training set is small, the feature distribution may only coarsely characterize feature space; in addition the a priori probability of seeing a particular object at random in the training set is also a highly unreliable estimate of seeing that object in the world, yet plays a powerful role in Bayesian decision functions [Duda73,Wood78]. Section 4.1.1 discusses in more detail the problems inherent in Bayesian methods when the statistical samples contain inaccurate or imprecise information, see also [Low82,Wes84].

2.1.2 The Plausibility Formalism

The approach used in our plausibility formalism differs in many aspects from the rule system. Whereas the high level goal of producing object hypotheses for a knowledge directed schema network is common to both systems, our approach is not thought of as producing a ranking of regions for each symbolic object (although the results may ultimately be used in this manner). The outcome is instead viewed as associating with each region an evidential model of the current state of interpretation.

Our general approach resembles the principles behind a least commitment or constraint propagation system. The system uses a set of features to rule out possible object hypotheses until only a small set of initial hypotheses remain. Each piece of evidence results in an assignment of a *plausibility value* to each hypothesis, yielding a *plausibility function* defined on the set of hypotheses. Each plausibility value is generated by comparing a feature value for an image event against the knowledge base. For example, the system may measure the average color of region 2; the value returned is then compared with the knowledge base for each of road, foliage, sky... to find a plausibility value for each possible semantic abstraction. The plausibility value represents the amount to which that hypothesis should not be ruled out as a possible hypothesis. Each plausibility function represents a mass function (a method for transforming a plausibility function into a mass function is discussed in Section 3.). The plausibility functions derived from each feature measurement are combined such that they represent the plausibilities of each singleton hypothesis as if the equivalent mass functions were combined using Dempster's Rule. The result is a combined plausibility function which represents a consensus of opinions from disparate pieces of evidence. Figure 2.2 shows the construction of a plausibility function given a feature measure and a knowledge base.

It is important to understand that the plausibility functions returned by each knowledge source are not statements about the *probability* of a hypothesis in a Bayesian view. The role of a plausibility value is to rule out unlikely states of nature. The minimum semantic requirement for a plausibility value is that it represent the extent to which a state of nature should not be ruled out given an event.

2.1.3 Differences in Evidential Representation

In the rule system, a vote can have one of three different effects on an object hypothesis. If the vote is between zero and one, this supplies supporting evidence, a vote of $-\infty$ vetoes

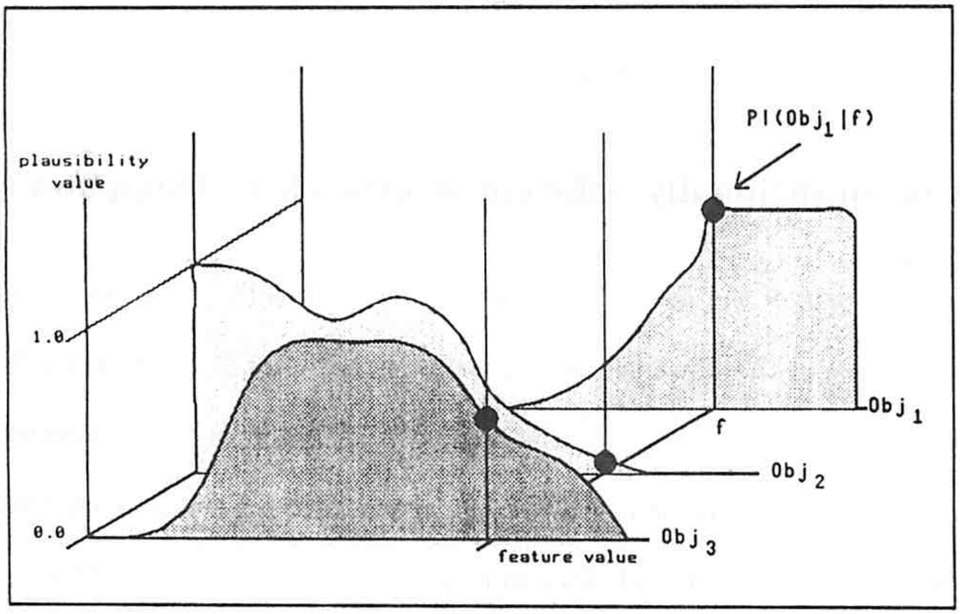

Figure 2.2: Constructing Plausibility Functions
The three shaded curves represent the knowledge base for Obj_1, Obj_2, and Obj_3 in the form of ***plausibility distributions*** for some feature. Given a feature value f, the plausibility value $Pl(Obj \mid f)$ can be determined from the plausibility distribution as shown by the shaded circles. A plausibility function is the set of plausibility values obtained for a given feature value.

an object hypothesis altogether, and a vote of zero is noncommittal. The final score represents supporting evidence for a particular object. On the surface, the plausibility formalism only allows ruling out hypotheses, but by transforming a plausibility function into a mass function after combination we are also able to represent supporting evidence as well as the possibility that an image abstraction may not be any of the objects represented by the set of possible hypotheses, i.e. conflicting evidence.

The greatest advantage of the Dempster-Shafer approach is its representation of several different types of evidence: supporting evidence, plausible evidence, and conflicting evidence. It greatest disadvantage is the exponential nature of Dempster's Rule and the explicit representation of the powerset of all object hypothesis. Our system provides a technique that allows the representative power of the Dempster-Shafer approach to be combined with the computational efficiency of the rule system.

3. A Computationally Efficient Approach to Dempster-Shafer

A major criticism of the Dempster-Shafer formalism has been the combinatorial problems related to use of the powerset of all possible hypotheses in Dempster's Rule. As the frame of discernment becomes large, the computation becomes unmanageable. To overcome this combinatorial problem, we present a way to represent a mass function using only the elements of the frame of discernment and a combination rule that is equivalent to Dempster's Rule for this simplified representation. Reynolds et al. [Rey86b] describe a method whereby knowledge sources need not return a mass function, but rather a *plausibility function* where each individual element of the frame of discernment, Θ, is assigned a *plausibility value* between zero and one. A mass function representation for a given plausibility function can be obtained using the formula:

$$m(A) = \frac{\prod_{a\in A} pl(a \mid f) \prod_{a\in\neg A} (1 - pl(a \mid f))}{1 - k}. \tag{3.1}$$

where $pl(a \mid f)$ is the plausibility of seeing object a given a feature value f (see figure 2.2). The conflict value of a plausibility function is defined as

$$k = \prod_{a\in A} (1 - pl(a \mid f)). \tag{3.2}$$

Given two plausibility functions $pl_1(a \mid f_1)$ and $pl_2(a \mid f_2), a \in \Theta$, we define the combination $pl_3(a \mid f_1 \wedge f_2)$ by the rule

$$pl_3(a \mid f_1 \wedge f_2) = \frac{pl_1(a \mid f_1) \cdot pl_2(a \mid f_2)}{1 - k}, a \in \Theta, \tag{3.3}$$

where k is the conflict value defined above for $pl_3(a \mid f_1 \wedge f_2)$ Analogously we define the combination of an arbitrary number of plausibility functions.

It is shown in [Rey86b] that this combination rule produces the the plausibilities of the singletons, as defined in the Dempster-Shafer theory, when the mass functions generated by formula 3.1 are combined using Dempster's Rule. The terms *plausibility* values and *plausibility* functions were chosen to point out this equivalence. A complete discussion and proof of the equivalence of this simple representation and multiplicative rule with a class of mass functions and Dempster's Rule can be found in Reynolds et al. [Rey86b].

4. Defining the Knowledge Base

In any image understanding scheme one of the most ill-defined tasks is that of defining the knowledge base. To overcome the problems of an inaccurate training set, the rule system creates its knowledge base by hand using the histogram of a set of training instances for heuristic guidance. On the other hand, Bayesian techniques address the problem of efficient definition of object rules for a given feature in terms of their conditional probability distribution, but these techniques lack the ability to handle inaccurate or insufficient sample sets. In this section a method is discussed for building a knowledge base, called *plausibility distributions*. This method constructs plausibility distributions determined directly by the statistics of a training set of image primitives (in this sense similar to Bayesian techniques), but can also deal with certain kinds of uncertainty inherent in the training set. In particular we address the situation where the training set is sufficient to show where an object lies in feature space, but the a priori estimate of seeing that object in the world, as estimated by the number of samples in the training set, is inadequate with regard to identifying objects in feature space.

4.1 Criteria for Plausibility Distributions

The approach taken by Hanson and Riseman [Hans86] describes a method for heuristically assigning a rule score based on the feature distributions for a particular object in relation the feature distribution for the entire image (or set of images). Our intent is to use statistical information to *automatically* constructs a knowledge base of plausibility distributions given information information about the frequency distribution of a set of training instances.

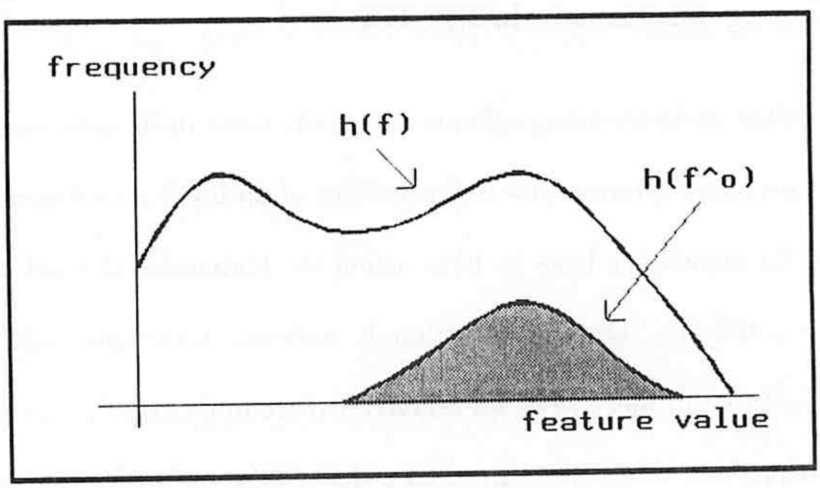

Figure 4.1: Frequency distributions

The larger curve represents the frequency of seeing a particular feature value, denoted by $h(f)$. This is the distribution of the feature over the entire population of samples. The smaller, shaded, curve represents the frequency of seeing a particular feature value and a particular object, $h(f \wedge o)$, the distribution of the feature over only those samples labeled object o.

The need is to find some function which characterizes the area in feature space in which an object lies and also takes into account the relationship between the object and total world sample. If we pick a set of pixels to represent some object, and a set of features over which to characterize each object, then we have the following statistical information: the frequency of seeing a particular feature value, $h(f)$ and the frequency of seeing a particular feature value and a particular object, $h(f \wedge o)$. (See Figure 4.1.)

Some considerations in evolving a system for producing automatic plausibility distributions include:

Statistical Accuracy: What statistical information is available from a training set of object instances, what assumptions can be made about the accuracy of these statistics, and how useful is this information for discrimination and recognition?

Semantics of a Plausibility Value: What are the specific semantics of a *plausibility value* with respect to our plausibility formalism.

Size of the Training Set: Should the final plausibility distribution be independent of the number of positive training instances for a particular object in the world? For example, if 30% of the training instances are $object_i$ and 20% are $object_j$, should the relative *plausibility* of finding $object_i$ over $object_j$ be dependent upon the probability of picking $object_i$ out of the training set at random (this is the classic discussion about the validity of a priori statistics)? It may or may not be reasonable to assume a priori that certain objects are more likely to be seen than others.

We briefly discuss each of these considerations in the next three sections.

4.1.1 Statistical Accuracy

We must first address the question of the accuracy of statistics collected over a set of training instances. Any matching or inference technique is only as good as the representation of the world knowledge as defined by a training set, and the quality of the information returned by the processes which measure the features to be compared with the world knowledge. In complex domains, both areas are subject to uncertain, imprecise and occasionally inaccurate information [Low82]. One concern with probabilistic methods is the amount of information needed to assure quality statistics. To accurately compute the conditional probability of seeing a particular state of nature ω given a feature measure,

$p(\omega \mid f)$, requires the a priori knowledge of seeing ω given no other information $p(\omega)$, the probability of perceiving a particular feature measure $p(f)$, and the probability of seeing a particular feature measure given a state of nature $p(f \mid \omega)$. At best, this information is difficult to obtain or reliably estimate. Sources of uncertainty can arise through inaccurate feature measures, incomplete data sets, aliasing resulting from digitization and region segmentation, and inaccurate region segmentation.

4.1.2 Semantics of a Plausibility Value

Secondly, we look at the desired semantics of a plausibility value. The objective at this stage of interpretation cannot be emphasized enough. When the interpretation begins, each token is possibly any object contained in the frame of discernment or the unknown event, moreover the features used to characterize object hypotheses at this level are extremely primitive. Therefore the goal is not object classification, but rather a pruning of possible hypotheses. With this in mind, the minimum semantic requirement of a plausibility value is that it represent how much an object should not be ruled out given a particular feature measure.

4.1.3 Size of the Training Set

Finally, some understanding must be reached about the size of the sample set for each object in the frame of discernment. Using Bayesian techniques, the conditional probability $p(\omega \mid f)$ represents both the probability of seeing object ω, $p(\omega)$, and knowledge about the probability of seeing feature measure f given ω, $p(f \mid \omega)$.

$$p(\omega \mid f) = \frac{p(f \mid \omega)p(\omega)}{p(f)}.$$

Consider the two object case, with objects ω_1 and ω_2. Given equal probability for $p(f \mid \omega_1)$ and $p(f \mid \omega_2)$ then the decision process is based solely on the the values for $p(\omega_1)$ and $p(\omega_2)$. In our paradigm of using sample regions hand labeled as the training set, then these two a priori probabilities are based solely on the number of samples for any object in the training set. This is not an adequate estimate. A compromise has been suggested that the a priori probability $p(\omega)$ is assumed equal for all objects and thus can be removed from the calculation leaving the likelihood ratio,

$$\frac{p(f \mid \omega)}{p(f)}.$$

A ratio greater than 1 indicates that the feature measure f is more likely to occur for this particular object than for some random object. A ratio equal to 1 indicates no information, and less than 1 states that the feature value f is more likely to occur by chance for any object than it is for a particular object ω_1. This ratio is the theoretical foundation underlying the heuristic approach of Hanson and Riseman [Hans86] and has also been employed by Woods in speech understanding [Wood78]. If the a priori statistics are assumed to be inaccurate, this is still not a reasonable solution because the a priori statistic $p(\omega)$ is still present in the term $p(f \mid \omega)$.

If it is reasonable to assume that certain object are more likely to appear than other object, then some estimate of the a priori probability may be devised. On the other hand, if you assume that the a priori probabilities are ill defined at best, then it is not good enough to simply remove that term from the computation of the conditional statistic, rather some effort must be made to factor out the size of the sample over which the probabilities are calculated.

In the domain considered here the number of samples for a given object, o, is not be an adequate estimate of the a priori probability of seeing that object in the world. Indeed this a priori probability may be impossible to determine. However the number of samples does influence $h(f)$ and $h(f \wedge o)$. What is needed is to find the range of feature values associated with a given object and to factor out any effects related to the probability of picking an object at random out of the training set.

In summary, early work on the generation of feature rules by Hanson and Riseman suggests that the rule for an object should be influenced not only by the its conditional feature distribution, but also by its relationship to the feature distribution of all the objects in the entire training set. In addition, we set as our goal that the final plausibility distribution should be characterizations of feature ranges and not include information about a priori probabilities.

4.2 Automatic Generation of Plausibility Distributions

For a given feature value consider the ratio of the height of a frequency distribution for a given object, $h(f \wedge o_i)$, to the height of the frequency distribution for all training objects, $h(f)$ (see Figure 4.1). This is by definition an estimate of the conditional probability [1]

$$\hat{p}(o_i \mid f) = \frac{h(f \wedge o_i)}{h(f)}. \tag{4.1}$$

Similarly defined are the estimates $\hat{p}(o_i)$, the relative frequency of seeing $object_i$ in the knowledge base and $\hat{p}(f)$, the relative frequency of seeing a particular feature value f.

$$\hat{p}(o_i) = \frac{\sharp of object_i}{\sum_{j=1}^{n} \sharp of object_j} = \frac{\sum_f h(f \wedge o_i)}{\sum_f h(f)} \tag{4.2}$$

[1] We will use the notation $\hat{p}$ to indicate the use of estimates of probabilities based on discrete samples.

$$\hat{p}(f) = \frac{h(f)}{\sum_f h(f)} \tag{4.3}$$

As mentioned earlier, one desirable characteristic of a plausibility distribution is that it be relatively invariant with respect to the size of the training set used to model the feature distribution. That is, two distributions differing only in the size of the sample set over which they are defined should have the same plausibility distribution. We can show that one way of accomplishing this behavior is to use the estimate of the a priori probability, $\hat{p}(o_i)$, as a decision threshold on the conditional probability $\hat{p}(o_i \mid f)$. If the value of $\hat{p}(o_i \mid f)$ is at least as large as the estimate of seeing o_i, $\hat{p}(o_i)$, assume a plausibility value of 1, otherwise normalize by $\hat{p}(o_i)$ We now have the following definition for a plausibility function.

Definition: *For each object o_i and each feature f we define a plausibility value for object o_i given feature f as follows:* [2]

$$Pl(o_i \mid f) = Min(1.0, \frac{\hat{p}(f \wedge o_i)}{\hat{p}(f)\hat{p}(o_i)}). \tag{4.4}$$

We can now examine the behavior of a plausibility value with respect to the size of the training set that makes up our statistical samples. To do this we must look at the behavior of

$$\frac{\hat{p}(f \wedge o_i)}{\hat{p}(f)\hat{p}(o_i)} \tag{4.5}$$

with respect to two objects with the same distribution but with difference sample sizes. Given two objects, o_1 and o_2, we can show that $Pl(o_1 \mid f) = Pl(o_2 \mid f)$ regardless of the difference in the size of $\hat{p}(o_1)$ and $\hat{p}(o_2)$ as long as the distribution for the two objects occupy the same area in feature space.

[2] A full discussion of this definition and proof of related theorems can be found in [Rey86b].

Theorem: *Let $\hat{p}(f \wedge o_2) = \alpha\hat{p}(f \wedge o_1)$ and $\hat{p}(o_2) = \alpha\hat{p}(o_1)$ where α is some scalar, then*

$$\frac{\hat{p}(f \wedge o_2)}{\hat{p}(f)\hat{p}(o_2)} = \frac{\alpha\hat{p}(f \wedge o_1)}{\hat{p}(f)\alpha\hat{p}(o_1)} = \frac{\hat{p}(f \wedge o_1)}{\hat{p}(f)\hat{p}(o_1)} \tag{4.6}$$

We show in figure (4.2) that this function indeed characterizes the proportion of feature space in which the objects feature distribution lies and minimizes the effect of the sample size over which the statistics are taken as stated by the above theorem.

5. Using Plausibility Functions for Initial Hypothesis

In this experiment we started with six images of New England road scenes digitized to a resolution of 256 x 256 pixels. Each image was then segmented using a knowledge based segmentation technique [Hans87]. Each region was hand labeled as one of the following labels {ambiguous, barn, dirt, foliage, grass, gravel, house, phonepole, pole, post, railing, road, roadline, sign, sky, sky-tree, tree-trunk, wire}, but only objects with a significant number of occurrences in the training set were used in the frame of discernment. The context for this experiment is defined by the following question, frame of discernment and set of feature spaces:

Question: "What are the plausible semantic labels for this region?"

Frame Of Discernment: {foliage, grass, gravel, road, roadline, sky, sky-tree, trunk}. [3]

Feature Spaces: Four feature categories were used; Intensity, Color, Location and Texture.

- **Intensity Features:** Y color transform, Intensity color transform, Raw red, Raw green, Raw blue.
- **Color Features:** Percentage of red, Percentage of green, Percentage of blue, Excess red, Excess green, Excess blue, Hue, Saturation, Q color transform, I color transform.

[3] The unknown object is not explicit but rather implicitly represented by the conflict value k.

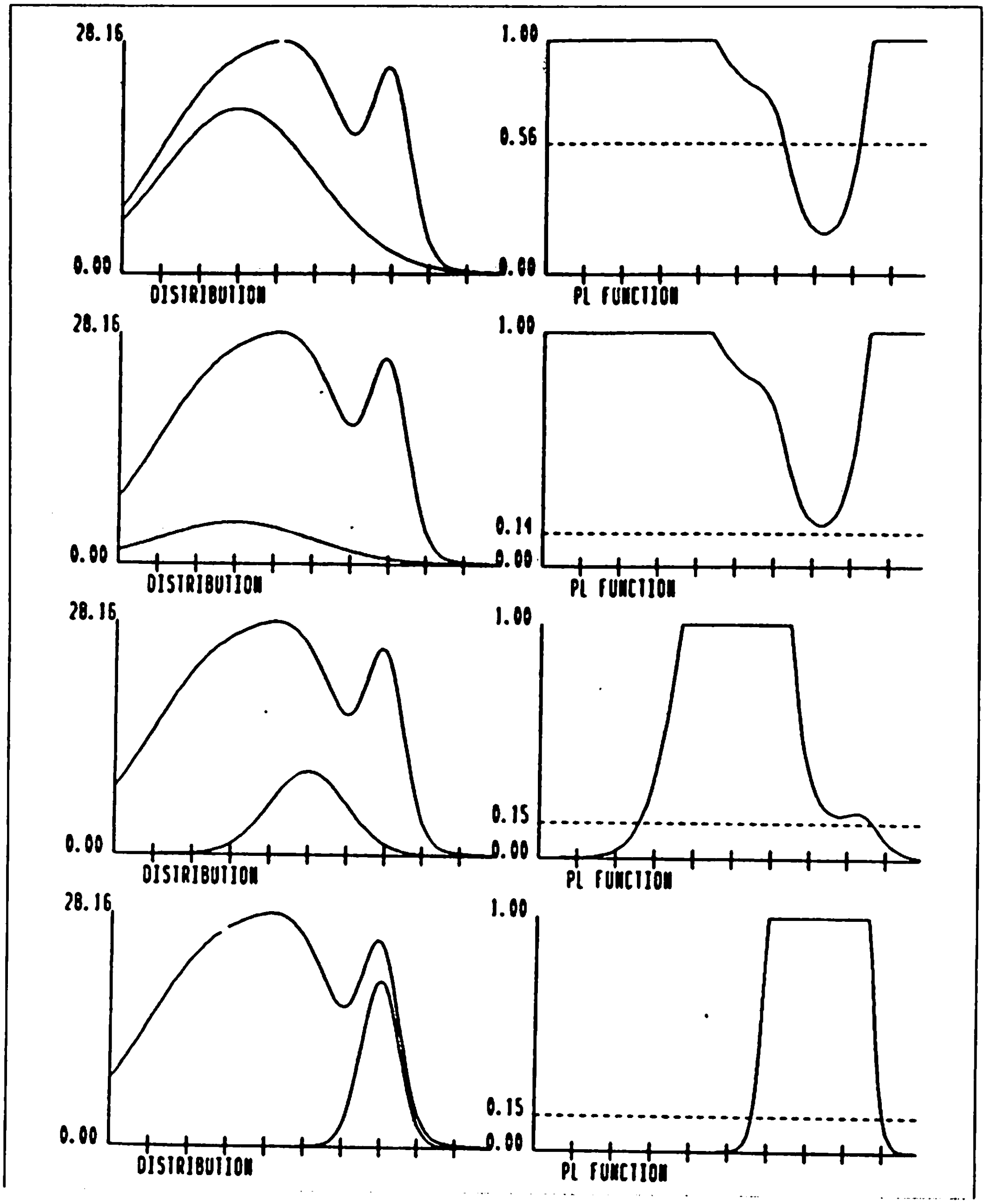

Figure 4.2: Plausibility distributions of synthetic world knowledge. This graph shows the plausibility distributions for synthetic world knowledge as defined by a set of four gaussian distributions. On the left is displayed $h(f)$, the sum of the gaussians, with $h(f \wedge o)$ shown separately from top to bottom. On the right, the four plausibility distributions are displayed. Note that the top two distributions have the same mean and standard deviation, and their plausibility distributions are identical. The dotted horizontal lines pass through $\hat{p}(o)$.

- **Texture Features:** Horizontal edge measure, Vertical edge measure, Lower diagonal edge measure, Upper diagonal edge measure.
- **Location Features:** Row position of centroid.

No information about object size or shape were used in this set of experiments. The objective was to use a simple set of feature measures to reduce the set of possible object hypotheses for any particular region. In particular, the feature measures used here contain little or no special knowledge which relates to a specific context. The approach is not strictly classification, but rather to provide some initial evidence for an hypothesis. The initial hypothesis can then provide information about a more specific context to be used as interpretation continues; with this more specific context are more specific, perhaps more expensive, feature measures. Under current development is a system for extending and verifying an initial hypothesis using a high-level knowledge based system implemented in a black board architecture [Weym86,Drap86].

The knowledge base for each feature space was formed over the feature distribution of the hand labeled regions from all six images. The feature spaces were specifically designed to use only features that could be measured over pixels. For each feature, a feature plane is defined which encodes a feature value at every point in the image. The feature planes and the training regions are then used to create a pixelwise frequency distribution for each object in the frame of discernment. An alternative is to define the frequency distributions by the mean feature value defined over the training regions and to weight the mean value by the size of the region. In the latter method, a high degree of smoothing is required to produce a robust frequency distribution. In the pixelwise method only minimal smoothing was used. During the interpretation phase, each knowledge source returns an evidential model based on the mean feature value of the region in question.

In addition, no plausibility was allowed to receive a value of zero. Due to the multiplicative nature of our combining function, a value of zero could cause a hypothesis to be completely ruled out based on errorful information. Instead, all plausibility values were normalized between .1 and 1. Figure 5.1 shows the plausibility functions generated for the feature intensity, where intensity is defined as $R + G + B/3$.

5.1 Understanding Normalization

In Section 3. we discuss a combination function for plausibility functions which parallels the Dempster's Rule applied to mass functions. This combination function uses $(1 - k)$ as a normalization constant. Intuitively, the amount to which two knowledge sources do not agree, k, indicates the amount to which the correct initial hypothesis is not contained in the frame of discernment. This section introduces other possible uses of normalization and different normalization constants when conflict is due to a situation other than the disagreement of knowledge sources.

5.1.1 Conflict Due to Inaccurate Evidence

An approach taken in the rule system is to load the system with redundant information so that no one knowledge source contributes a significant amount of information to the interpretation process. This is desirable if any of the knowledge sources are suspected to be unpredictably errorful. Unfortunately, in the approach presented here, as the number of objects and feature spaces increase, so does the conflict between plausibility functions in the interpretation of a token. In particular the automatic plausibility distributions may allow many objects to receive a small plausibility value, ruling out completely no one object

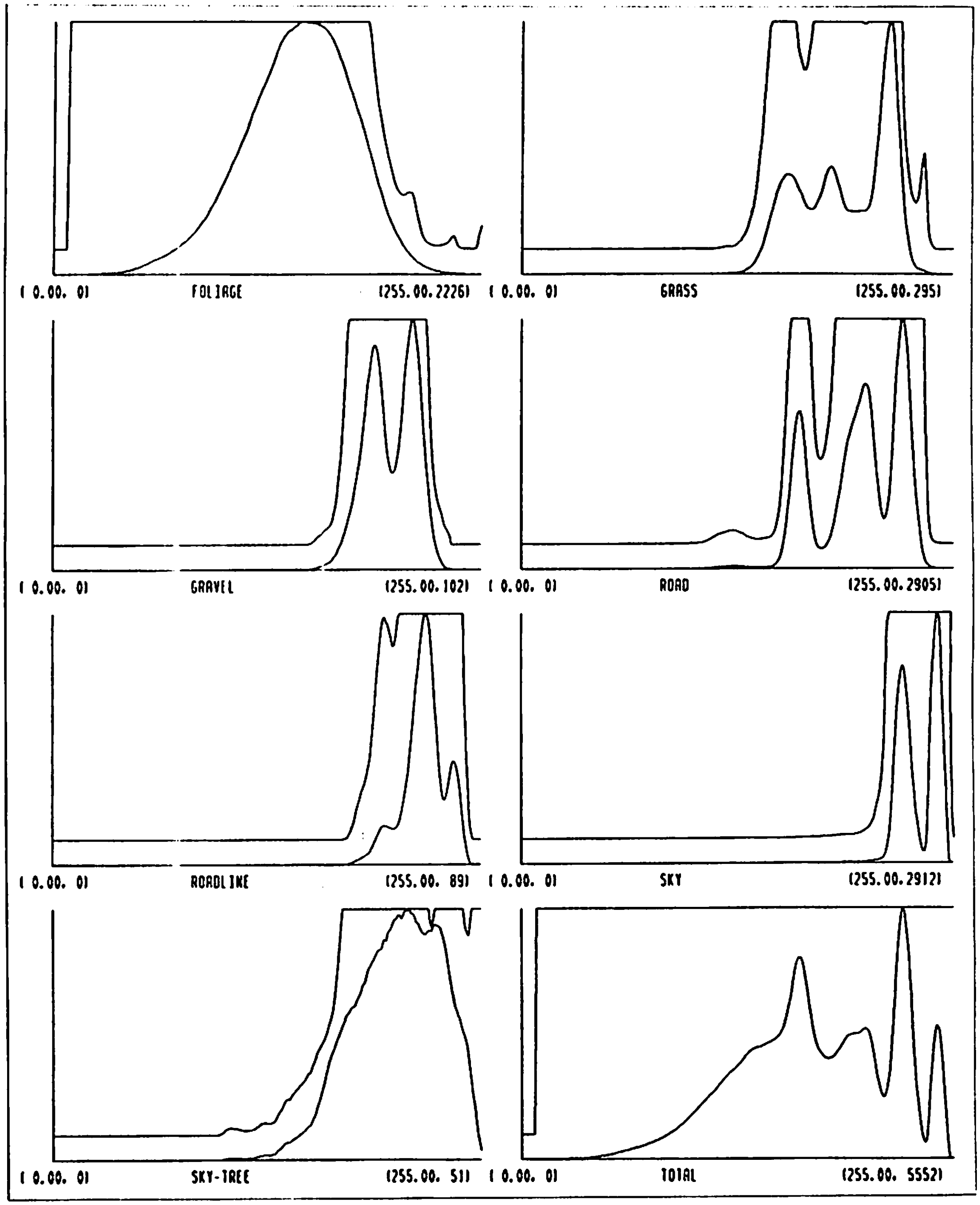

Figure 5.1: Plausibility distributions of Intensity.

These graphs show the feature histograms and resulting plausibility functions for the objects foliage, grass, gravel, road, roadline, sky, sky-tree, and total. For each object, the lower curve is the histogram of the actual feature values for the regions in the training set. The upper curve is the resulting plausibility function. Each object histogram is scaled with respect to the number of samples contained in the training set for that particular object. The histogram for "total" in the lower right hand corner represents all regions in the training set including those whose labels are not included in the set of possible initial hypotheses.

for any one feature value. For many tokens, each object in the frame of discernment will be ruled out to some significant degree by at least one piece of evidence due to errorful data (e.g. inaccurate feature measurement due to aliasing or poor placement of a segmentation boundary). This situation introduces conflict into the final interpretation which is not due to the detection of an unknown image event.

Recognizing that conflict can be generated by a single knowledge source, it may be reasonable to eliminate this conflict before combination with other knowledge sources. This can be accomplished by normalizing a plausibility function by its maximum plausibility value.

$$pl_{norm}(a \mid f) = \frac{pl(a \mid f)}{MAX_{a \in \Theta} pl(a \mid f)}$$

5.1.2 The Associative Issue

We have stated an equivalence between plausibility functions and mass functions. If n plausibility functions are multiplied term-wise and normalized by the conflict value of the combined plausibility function, then the result is the plausibilities of the singletons obtained by combining a set of n equivalent mass functions using Dempster's Rule. What happens when a new piece of evidence becomes available? Dempster's Rule is both commutative and associative. This says that new evidence represented as a mass function can be combined with evidence contained in a previously combined mass function with the same effect as having the new knowledge available from the beginning of the inference process. In the case of the plausibility functions this is not true.

Let *cpl* be the result of combining n plausibility functions and *cm* the result from n mass functions. Although *cpl* are the plausibilities of *cm* over the singleton sets of the frame of discernment, the mass function obtained from *cpl* by formula (3.1) is not the same mass function as *cm*. Therefore the addition of a new plausibility function to *cpl* will not yield the same plausibilities as the addition of a new mass function to *cm*. The discrepancy is caused by the fact the the combined plausibility functions are normalized only once, after all the term-wise multiplications have been performed, conversely Dempster's Rule normalizes after each pair wise combination. A simple solution is to represent an evidential pair consisting of a plausibility function and the conflict value by which it was normalized. Now we can construct a combination rule that is equivalent of Dempster's Rule over this pair.

$$pl_3 = E(pl_1, k_1) \bigoplus E(pl_2, k_2)(o_i) = \frac{(1-k_1)pl_1(o_i) * (1-k_2)pl_2(o_i)}{1-k_3} \tag{5.1}$$

Part of the Dempster-Shafer approach assumes conditionally independent knowledge sources, that is knowledge from one piece of evidence does not provide any information about another piece of evidence. When this assumption is violated by multiple redundant knowledge sources we must find a way to realize independence. The approach taken in the Rule System is to group redundant knowledge sources into independent groups, combine the groups of knowledge sources and then combine the results of each group. This will have an effect on the final combined evidence if and only if the combination rule is *not* associative. When dealing with redundant multiple knowledge sources, the non-associative nature of combining plausibility functions can be an advantage.

5.1.3 In Defense of Normalization

It has been argued that normalization is used to eliminate or *hide* a contradiction of aggregate evidence [Zad83]. It must be noted that with a multiplicative combination function, the use of no normalization will create monotonically decreasing plausibility values. This means that any evidence supplying a plausibility value less than 1.0 can only decrease a hypothesis' plausibility value. The more evidence that is accumulated, the more the system becomes susceptible to errorful data.

In the results presented in this paper, we have chosen to use a normalization constant of $(1 - k)$. Using this normalization keeps the system from being a monotonically decreasing system, preserves the semantics of the plausibility values as the plausibilities of the singleton subsets of the related mass function, and allows for the representation of an unknown object.

5.2 Interpretation Strategies and Future Work

In these experiments, the Dempster-Shafer formalism is not used as an end to reasoning, but rather as a representation formalism for evidence. The interpretation strategy to be employed after initial hypothesis generation dictates some of the desired qualities for a decision over the set of initial hypothesis. One approach might be to use the initial hypothesis to indicate exemplar regions, regions which most look like grass for example. These regions can then be used to constrain the interpretation of neighboring regions, on the other hand, the initial hypothesis can be viewed as simply reducing the set of possible hypothesis.

Several different decision strategies were suggested creating the initial set of hypotheses for this experiment. The simplest strategy is to normalize out all conflict contained in a plausibility function (by dividing the plausibility function by the maximum plausibility value) and use these numbers for ranking objects. This strategy does not allow the representation of the unknown object. Normalizing by $(1 - k)$ allowed an explicit representation of the unknown event. Another strategy suggested uses the idea of consistent evidence. A simple strategy can be used to find the most likely object. Next, all plausibility functions which rule out to some significant degree the top ranking object, as produced by the simple strategy, are then removed from the final combination of evidence.

The decisive factor for determining the best decision strategy at any step in the interpretation should be the context and objectives of the interpretation system as a whole. There may be only a few relevant control (control about conflict) strategies, or control may become a highly variant context dependent process.

6. Results and Figures

The hypotheses displayed in this section are initial hypotheses to be used by a knowledge-directed, opportunistic image interpretation system. The initial hypotheses can verified and used to create a contextual environment in which to hypothesize objects not defined by the frame of discernment, hallucinate the merging or splitting of image tokens, and in general aid in the development of a mapping between image events and world events.

We chose the following two step combination scheme using the multiplicative combination function defined by formula 3.3:

1. Combine all the plausibility functions for each of the four feature spaces (i.e. intensity, color, texture and location) and normalize each by $(1 - k)$, with k defined by for each plausibility function.

2. Combine the four resulting plausibility functions and again normalize by $(1 - k)$, with k defined by the combined plausibility function.

By combining the evidence for each category of features first, we were able let each group have equal weight in the final combination. For instance, there are ten color features, but only one location feature. By combining the color features and renormalizing before combining with the location feature, we were able to represent color and location equally in the final interpretation. Once the evidence contained in each plausibility function is combined, a mass function is constructed as discussed in Section 3. to represent the final evidential model.

Figure 6.2 through figure 6.5 each show four pieces of information for a particular object hypothesis. The terms, *support, plausibility, singleton objects,* and *mass value* are used in terms of the Dempster-Shafer theory of evidence. Brief definitions of these terms can be found in Appendix A. For each piece of information, the intensity encodes a numeric value with darker regions representing higher numbers.

- The top left quadrant displays all the regions which would answer the question "What are all the regions initially hypothesized as object X?", where X represents the object under consideration. An object is considered as an initial hypothesis if it is contained in the highest ranking subset of the final mass function.
- The top right quadrant shows the mass value assigned to the highest ranked subset for those regions displayed in the top left quadrant.
- The bottom left shows, for every region in the image, the support value for the set containing the singleton object under consideration. The support is computed from the final mass function and can be thought of as the lower bound on the system's belief that this region is this object.
- Finally, the bottom right displays the plausibility value for the set containing the singleton object under consideration. Again, the plausibility is computed from the final mass function and can be thought of as the upper bound on the system's belief that this region is this object.

In figure 6.7, the combined conflict obtained for each region is displayed. A high conflict value (i.e. a dark region) indicates a large amount of disagreement between separate pieces of evidence. This is the value for k contained in the normalization constant of the combination function (see formula 3.2). The conflict measure can be used to pinpoint areas which may need more verification. Also of interest is the conflict obtained by each intermediate combination. This is displayed from left to right, top to bottom as conflict from color, intensity, texture and location respectively in figure 6.8.

7. Conclusion

We have presented here a method which automatically generates a knowledge base for the formation of initial object hypotheses using statistical information provided from a set of training objects. The plausibility formalism uses a computationally efficient approach to the Dempster-Shafer formalism for representing and combining evidence. The Dempster-Shafer formalism is used for its rich evidential representation. Our system addresses the problems of using heuristic methods for constructing a knowledge base and combining evidence as well as the problems of a strict Bayesian approach. Presented is a set of results showing our plausibility theory applied to a set of color outdoor road scenes which shows that the approach has significant potential.

8. Acknowledgments

We would to thank Edward Riseman and Allen Hanson for their guidance and careful reading of this paper. In addition thanks go to Deborah Strahman for sharing her breadth of knowledge in the areas of probable, possible, and plausible reasoning.

Figure 6.1: Intensity Image

The intensity image digitized to a resolution of 256 by 256. The intensity was computed from the red, green, and blue planes using the formula $R + G + B/3$.

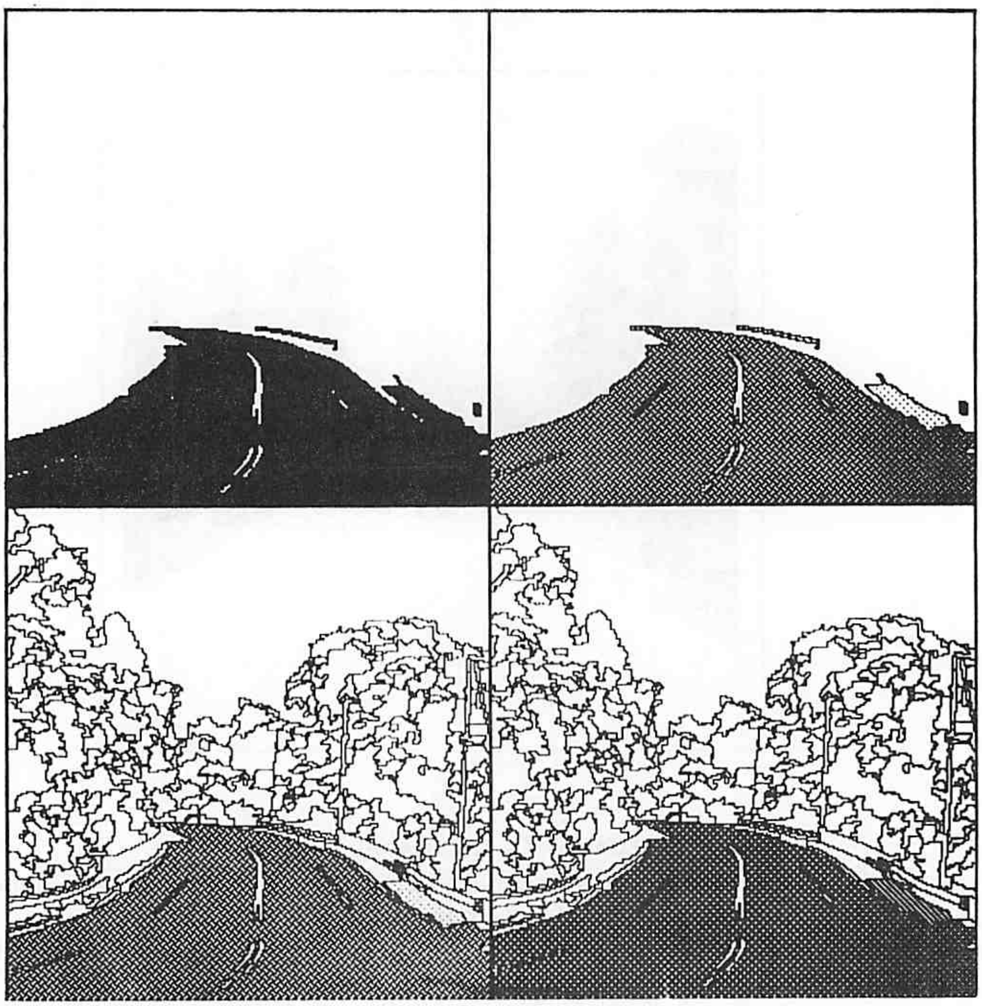

Figure 6.2: Initial Hypotheses for Road

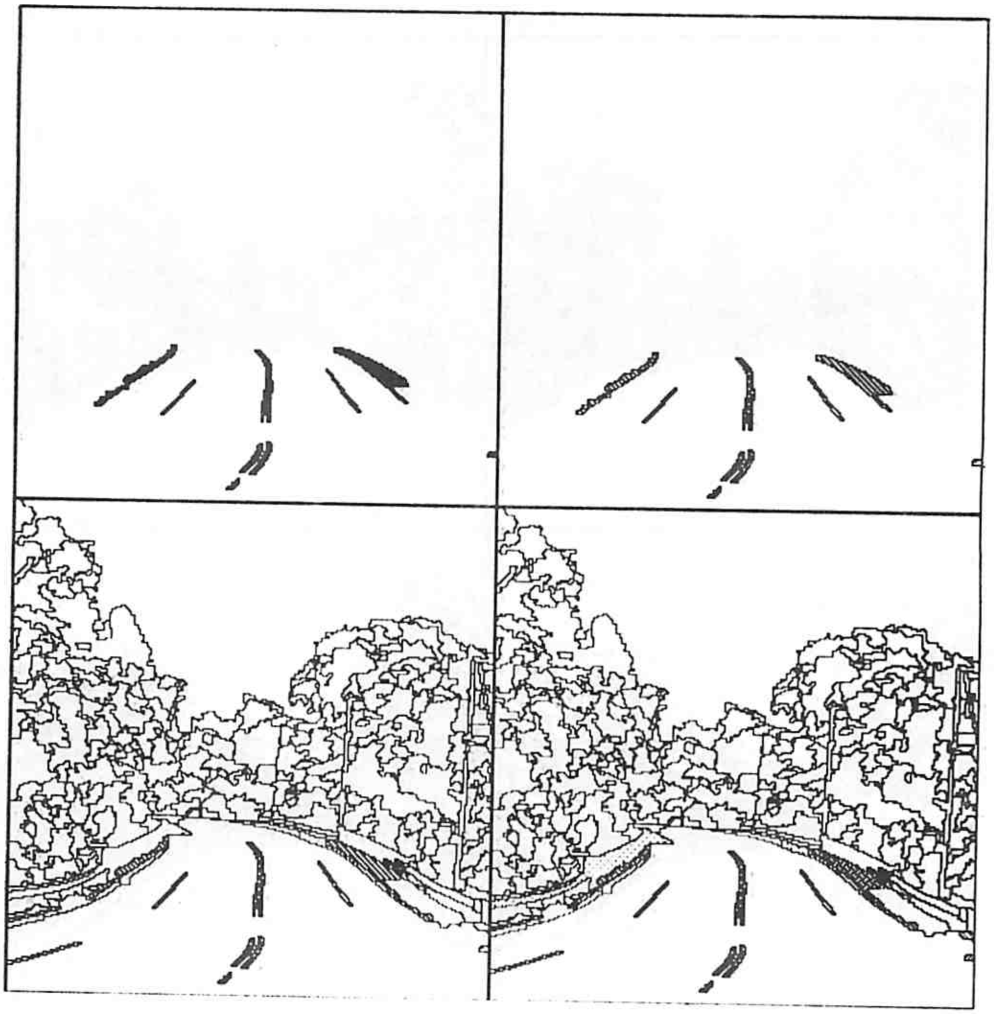

Figure 6.3: Initial Hypotheses for Road Line

Figure 6.4: Initial Hypotheses for Foliage

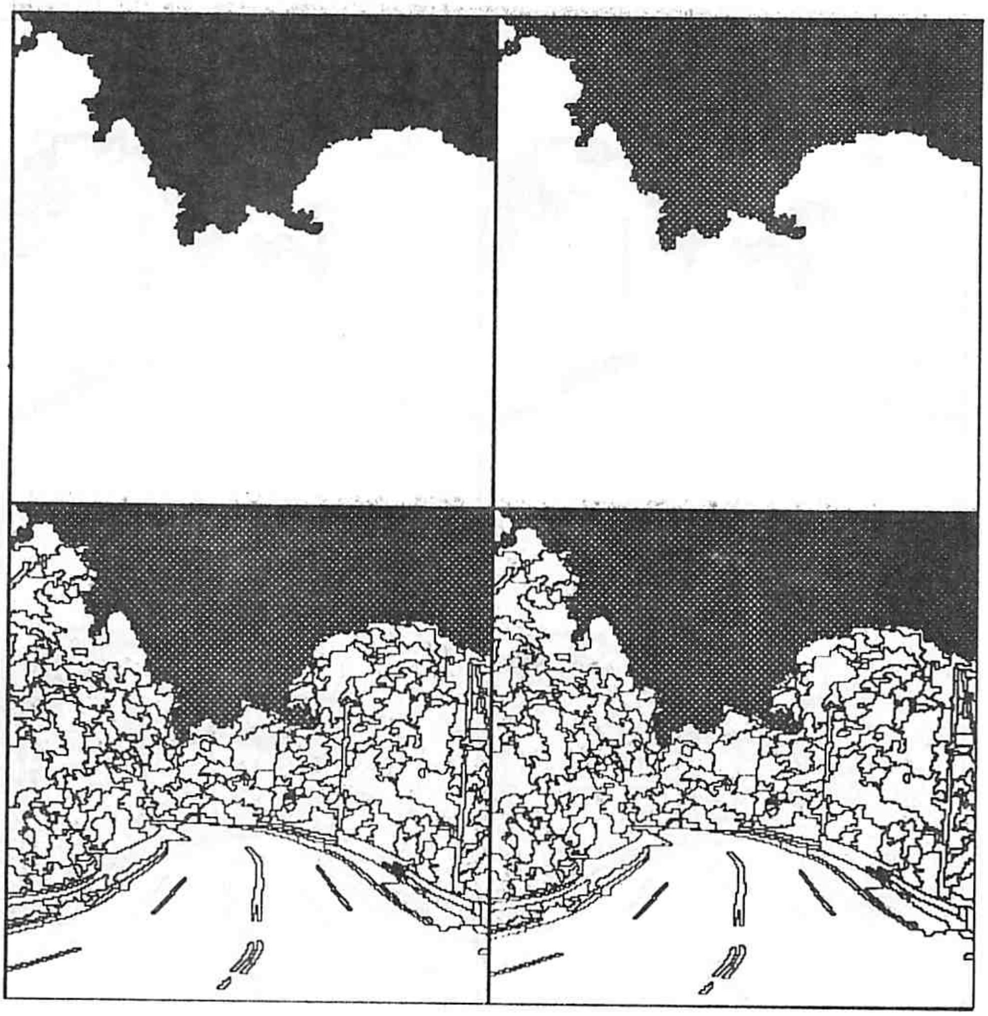

Figure 6.5: Initial Hypotheses for Sky

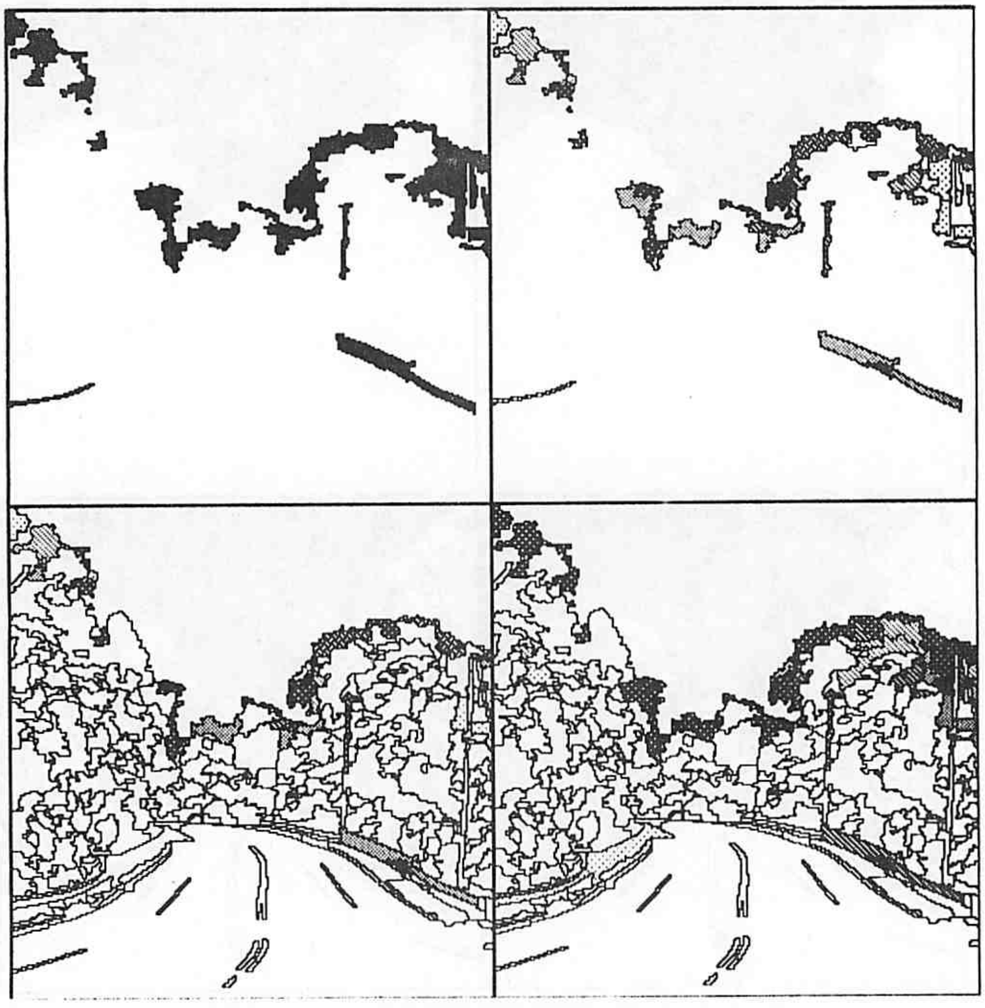

Figure 6.6: Initial Hypotheses for Sky-Tree

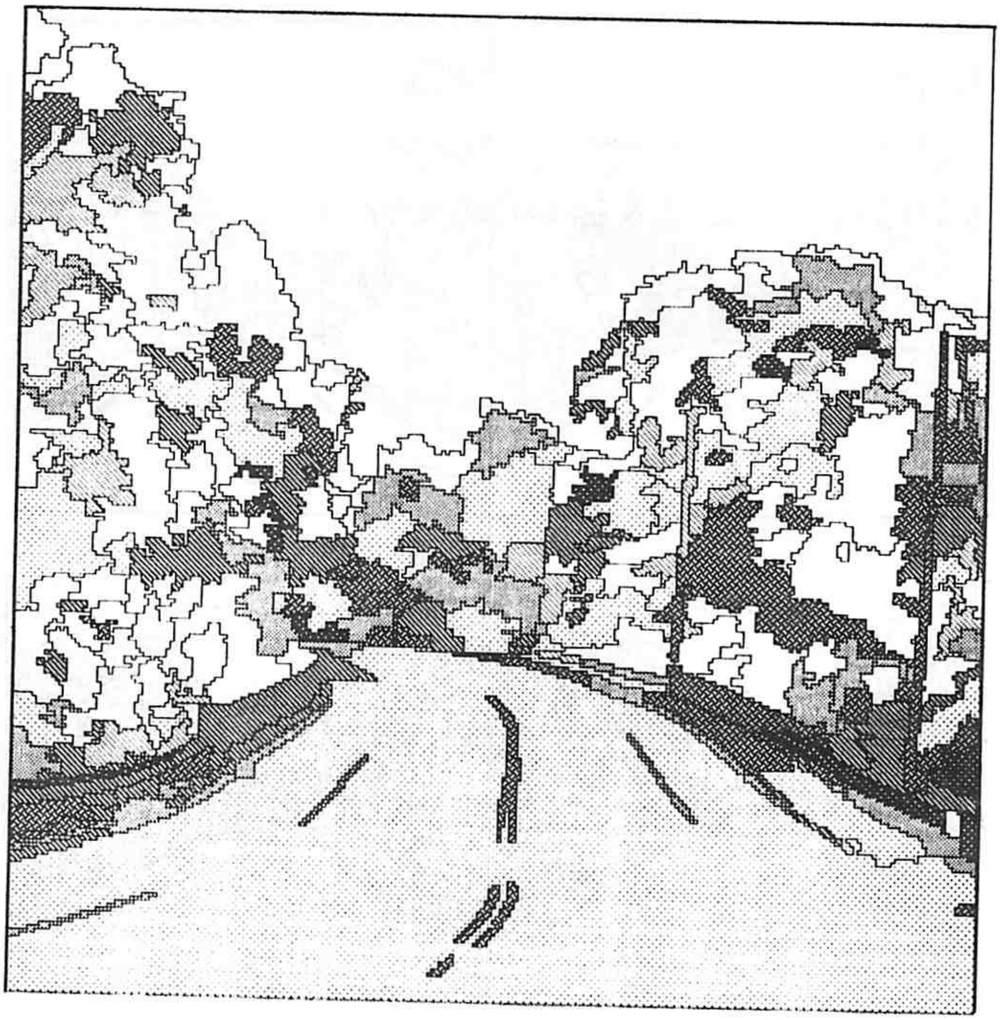

Figure 6.7: Total Combined Conflict

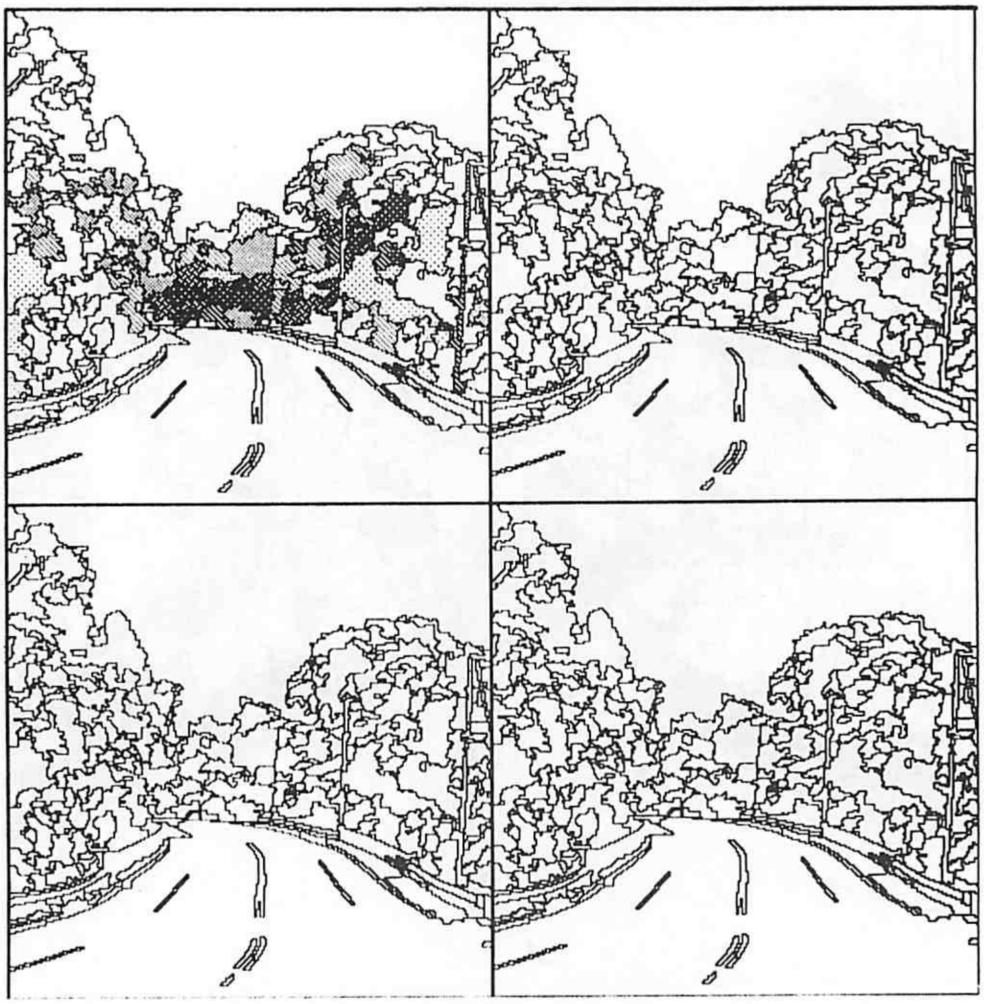

Figure 6.8: Conflict for Color, Intensity, Texture, and Location

A. Dempster-Shafer Tutorial

A.1 Dempster-Shafer — An Intuitive Approach

The Dempster-Shafer theory of evidence provides a rich set of semantics for representing and combining evidence and can be viewed as a least commitment approach. Each piece of evidence is intended to reduce a set of possible hypotheses. The consensus of many pieces of evidence is then used to focus on the smallest possible set of hypotheses. Dempster's combination rule is the heart of the system and much literature can be found describing its focusing methods [Shaf76,Lu84,Low82,Wes84,Rey86a,Rey86b].

A.2 Terms and Concepts

The Dempster-Shafer theory of evidence can be broken up into several concepts:

Frame Of Discernment: A framework which defines a set of possible hypotheses, or possible answers to a given question. The frame of discernment defines a set of assertions: *This is an* X. where X is an element of the frame of discernment. A frame of discernment is defined to contain all possible answers to the question to be answered. This assumption can be fulfilled by the addition of some "unknown" object which represents all answers not explicitly contained in the frame of discernment. The set which contains all elements of the frame of discernment is represented by Θ.

A Mass Function: A structure for representing evidence which maps a unit of belief over the powerset of the frame of discernment. The mass value assigned to a subset of the frame of discernment quantifies how much the belief is contained in the system that *exactly* that subset of hypotheses is the best answer to the question.

A Feature Space: Each feature space is fashioned to provide information needed to answer some particular aspect of the given question. A set of feature spaces are designed to bring together many aspects of the evidence needed for an inference.

A Knowledge Source: Maps a feature value for a specified feature space into a representation suitable for combining with other evidence given a combination rule. In this case, a knowledge source maps a feature value into a mass function to be combined with other pieces of evidence using Dempster's Rule.

Dempster's Rule: Dempster's Rule combines evidence as represented by two mass functions to create a consensus of opinion. Dempster's Rule is both associative and commutative providing means for several pieces of information to brought together as combined evidence. Mathematically Dempster's Rule is the orthogonal sum of two mass functions and is defined as follows: For every $C \in \theta$

$$m_1 \bigoplus m_2(C) = \frac{\sum_{A \cap B = C} m_1(A) \cdot m_2(B)}{1-k} \tag{A.1}$$

where

$$k = \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B). \tag{A.2}$$

k is precisely a statement about how much disagreement exits between the two pieces of evidence.

Support, Plausibility, Doubt, and Conflict: Metrics which describe the degree to which some member or some subset of the frame of discernment is the correct answer to a given question.

In the Dempster-Shafer approach, the object space is defined over the powerset of all possible objects. A hypothesis then is not merely an element of the frame of discernment, but rather a set of objects. Each knowledge source is responsible for distributing a unit of belief across the subsets of the frame of discernment. The resulting list of object subsets and mass values is a *mass function*. Higher mass values are associated with more likely hypotheses. Three restrictions apply to mass functions; each subset receives a mass value between zero and one, the empty set receives zero mass, and the sum of the mass values over the object space equals one. For a complete discussion of mass functions see [Shafer].

In general, an inference problem consists of a question to be answered, a frame of discernment containing all possible answers to this question, and a set of feature spaces each designed to provided some piece of information useful in answering the question at hand. This collection is termed a *context*.

A.3 Specific attributes of the Shafer Dempster Approach

As well as providing a combination rule for bringing together disparate sources of evidence, the Dempster-Shafer approach provides explicit information about supporting evidence, plausible evidence, uncertainty of a decision, conflict between knowledge sources, disbelief and no belief. The following subsections describe these types of information.

A.3.1 Supports, Doubt, and Plausibility

The mass values associated with each subset in the object space is only a small portion of the information supplied by a mass function. Of initial importance is the information obtained by asking the following questions for a given subset A: "What amount of of evidence supports A as being the correct answer?", "What amount of evidence refutes A as being the correct answer?" and "What amount of evidence fails to refute A as being the correct answer?". The values obtained by asking these questions are the referred to as *support*, *doubt* and *plausibility* respectively.

$$Spt(A) = \sum_{s \subseteq A} m(s).$$

$$Dbt(A) = \sum_{s \cap A = \emptyset} m(s).$$

$$Pls(A) = \sum_{s \cap A \neq \emptyset} m(s) = 1 - Spt(\neg A).$$

Figure A.1 shows graphically the relationship between the support, doubt and plausibility.

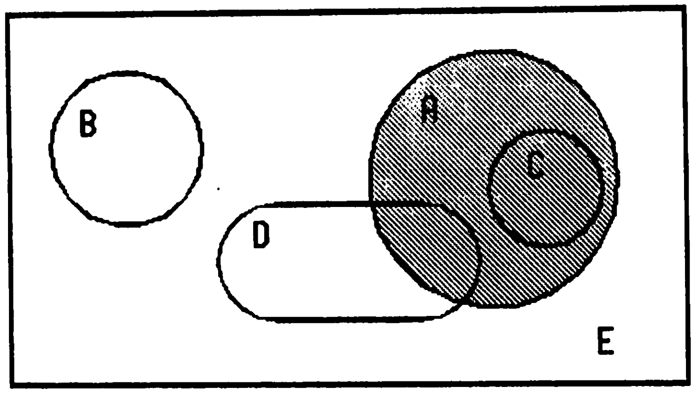

Figure A.1: Venn diagram of support, plausibility and doubt.
$Spl(A) = m(A) + m(C)$.
$Pls(A) = m(A) + m(C) + m(D) + m(E)$.
$Dbl(A) = m(B)$.

A.3.2 Knowledge about Conflict

Implicit with each combined mass function is a statement about how the knowledge sources involved in the decision process agree with one another. This is known as the conflict value and is precisely defined by Dempster's Rule as:

$$k = \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B). \tag{A.3}$$

One assumption mentioned earlier is that all possible objects are represented in the frame of discernment. This may be accomplished by the addition of the *unknown* object in the frame of discernment. If all knowledge sources are required to give some mass to the *unknown* object, then $k = 0$ for all combinations of mass functions and the conflict value is precisely the amount of mass assigned to this *unknown* object.

A.3.3 Disbelief versus No Belief

Disbelief and no belief can be discussed both in terms of a mass function as a whole and about a particular subset receiving mass within a mass function. Looking at a mass function as a whole, statements about support and plausibility are direct statements about disbelief and no belief. No belief is represented as the support for some subset being near zero, where as disbelief is represented as a plausibility near zero. Looking at the mass values themselves. If for a given subset A the mass value is near one, then for any $a \subseteq A$ the system provides little information (no belief) for the support of a but much information about its disbelief for any $b \notin A$. Disbelief about the systems performance in general (or about a knowledge source) is contained in the appropriate conflict value.

Bibliography

[Binf82] T. Binford, *Survey of Model Based Image Analysis Systems*, International Journal of Robotics Research 1 (1982), pp. 18-64.

[Belk86] R. Belknap, E. Riseman and A. Hanson (1985), "The Information Fusion Problem and Rule-Based Hypotheses Applied to Complex Aggretations of Image Events." *Proceedings: DARPA Image Understanding Workshop*, December 1985, pp. 279-292.

[Burn86] J. Brian Burns, Allen R. Hanson & Edward M. Riseman. "Extracting Straight Lines", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8, #4 (July 1986), pp. 425-455.

[Demp67] A. P. Dempster (1967), "Upper and lower probabilities induced by a multivalued mapping", *Annals of Mathematical Statistics*, vol. 38, 1967, pp. 325-339.

[Demp68] A. P. Dempster (1968), "A generalization of Bayesian inference", *Journal of the Royal Statistical Society*, Series B, vol. 30, 1968, pp. 205-247.

[Demp84] A. P. Dempster and A. Kong (1984), "Belief unctions and communications networks", Report, Department of Statistics, Harvard University, Cambridge, MA, Nov. 1984.

[Drap86] Bruce A. Draper, Allen R. Hanson & Edward M. Riseman (1986), "A Software Environment for High Level Vision". *Proceedings AAAI Workshop on Knowledge Based Systems and Tools*, Columbus, Ohio; October 1986.

[Duda73] R. Duda and P. Hart (1973), *Pattern Classification and Scene Analysis*. A Wiley-Interscience Publication, New York. 1973.

[Faug82] O. D. Faugeras (1982), "Relaxation labeling and evidence gathering", *Proc. Conf. Pattern Recognition and Image Processing*, Las Vegas, June 1982, pp. 672-677.

[Hans78] Hanson, Allen R. & Riseman, Edward M. "VISIONS: A Computer System for Interpreting Scenes", in *Computer Vision Systems*, Hanson & Riseman, eds., Academic Press, N.Y., 1978. pp. 303-333.

[Hans86] Hanson, Allen R. & Riseman, Edward M. (1986), "A methodology for the development of general knowledge-based vision systems", in *Vision, Brain, and Cooperative Computation* (M. Arbib and A. Hanson Eds.) to be published by MIT Press 1987: Also COINS Technical Report 86-27, University of Massachusetts at Amherst, July 1986

[Hans87] Hanson, Allen R. & Riseman, Edward M., "The VISIONS Image Understanding System – 86", in *Advances in Computer Vision*, C. Brown (Ed.), Erlbaum, 1987.

[Humm83] R. A. Hummel and S. W. Zucker (1983), "On the foundations of relaxation labeling processes", *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-5, no. 3, May 1983, pp. 267-286.

[Kitc80] L. J. Kitchen (1980), "Relaxation applied to matching quantitative relational structures", *IEEE Trans. Syst. Man Cybern.*, vol. SMC-10, 1980, pp. 96-101.

[Kitc84] L. J. Kitchen and A. Rosenfeld (1984), "Scene analysis using region-based constraint filtering", *Pattern Recognition*, vol. 17, no. 2, 1984, pp. 189-203.

[Kohl84] Ralf R. Kohler. *Integrating Non-Semantic Knowledge into Image Segmentation Processes.* Ph.D. thesis, COINS, Univ. of Massachusetts, Amherst, 1984.

[Kybu84] H. E. Kyburg (1984), "Bayesian and non-Bayesian evidential updating", Tech. Report 139, Department of Computer Science, University of Rochester, Rochester NY, July 1984.

[Low82] J. D. Lowrance (1982), "Dependency-graph models of evidential support", Ph.D. Thesis, University of Massachusetts, Amherst, 1982.

[Low83] J. D. Lowrance, T. D. Garvey (1983), "Evidential reasoning: an implementation for multisensor integration", Tech Note 307, SRI Artificial Intelligence Center, December 1983.

[Lu84] S. Y. Lu, H. E. Stephanou (1984), "A set-theoretic framework for the processing of uncertain knowledge", AAAI-84, pp. 216-221.

[Rey86a] G. Reynolds, D. Strahman and N. Lehrer (1985), "Converting Feature Values to Evidence", *Proceedings: DARPA Image Understanding Workshop*, December 1985, pp. 331-339.

[Rey86b] G. Reynolds, D. Strahman, N. Lehrer and L. Kitchen (1986), "Plausible Reasoning and the Theory of Evidence." University of Massachusetts at Amherst COINS Technical Report 86-11, April 1986.

[Ros76] A. Rosenfeld, R. A. Hummel and S. W. Zucker (1976), "Scene labeling by relaxation operations", *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, 1976, pp. 420-433.

[Shaf73] G. Shafer (1973), "A theory of statistical evidence", pp. 365-434 in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, W. L. Harper and C. A. Hooker eds, vol. II.

[Shaf76] G. Shafer (1976), *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[Shaf81] G. Shafer (1981), "Constructive Probability", *Synthese*, vol. 48, 1981, p1-60.

[Shaf84] G. Shafer (1984), "Probability judgement in artificial intellegence and expert systems", Working Paper 165, School of Business, The University of Kansas, Lawrence, Kansas, Dec. 1984.

[Shaf85] G. Shafer (1985), "Belief functions and possibility measures", *The Analysis of Fuzzy Information*, vol. 1, J. C. Bezdek, ed., CRC Press.

[Shaf83] G. Shafer and A. Tversky (1983), "Weighing evidence: the design and comparison of probability thought experiments", *75th Anniversary Colloquium Series*, Harvard Business School.

[Shor76] Shortliffe, E. H. (1976) *Computer-Based Medical Consultations: MYCIN*. New York: American Elsevier.

[Smit61] C. A. B. Smith (1961), "Consistency in statisical inference and decision", *Journal of Royal Statisical Society*, Series B, vol 23, 1961, pp. 1-37.

[Smit65] C. A. B. Smith (1965), "Personal probability and statistical analysis", *Journal of the Royal Statisical Society*, Series B, vol. 128, 1965, pp. 469-499.

[Strt84] T. Strat (1984), "Continuous belief functions for evidential reasoning", AAAI-84, pp. 308-313.

[Wal72] Waltz, D. (1972) "Generating Semantic Descriptions from Drawings of Scenes with Shadows." AI-TR-271. Project MAC, Massachusetts Insititute of Technology. (Reprinted in P. Winston (Ed.). (1975) *The Psycology of Computer Vision*. McGraw-Hill, New York, 19-92.)

[Weis86] Weiss, R. and M. Boldt. (1986) "Geometric Grouping Applied to Straight Lines." Computer Vision and Pattern Recognition, 1986.

[Wes83] L. Wesley (1983), "Reasoning about control: the investigation of an evidential approach", *Proc. Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, August 1983, pp. 203-210.

[Wes84] L. Wesley (1984), "Reasoning about control: an evidential approach", Tech. Report 324, SRI Artificial Intelligence Center, 1984.

[Wes85] L. Wesley (1985), Ph.D. Thesis, University of Massachusetts, Amherst, *in preparation*.

[Weym86] Terry E. Weymouth. *Using Object Descriptions in a Schema Network for Machine Vision*, Technical Report 86-24, COINS dept., Univ. of Massachusetts, May 1986.

[Wood78] Woods, W. A. "Theory Formation and Control in a Speech Understanding System with Extrapolations Toward Vision", in *Computer Visions Systems*, Hanson and Riseman, eds., Academic Press, N.Y., 1978. pp 379-390.

[Zad83] L. A. Zadeh (1983), "The role of fuzzy logic in the management of uncertainty in expert systems", *Fuzzy Sets and Systems*, vol. 11, 1983, pp. 199-227.