SRI International

EVIDENTIAL KNOWLEDGE-BASED COMPUTER VISION

Technical Note No. 374

21 January 1986

By: Leonard P. Wesley, Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
21 JAN 1986		**00-01-1986 to 00-01-1986**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Evidential Knowledge-Based Computer Vision	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**64**	

ABSTRACT

It has been argued that knowledge-based systems (KBS) must reason from evidential information – i.e., information that is to some degree *uncertain*, *imprecise*, and occasionally *inaccurate*. This is no less true of KBS that operate in the domain of computer-based image interpretation. Recent research has suggested that the work of Dempster and Shafer (DS) provides a viable alternative to Bayesian-based techniques for reasoning from evidential information. In this paper, we discuss some of the differences between the DS theory and some popular Bayesian-based approaches to effecting the reasoning task. We then discuss some work on integrating the DS theory into a knowledge-based high-level computer vision system in order to examine various aspects of this new technology that have not been explored to date. Results from a large number of image interpretation experiments will be presented. These results suggest that a KBS's performance improves substantially when it exploits various features of the DS theory that are not readily available in pure Bayesian-based approaches.

Index Terms: uncertain reasoning, evidential reasoning, belief functions, knowledge-based system, computer vision, image interpretation.

1 INTRODUCTION

It is widely accepted that knowledge based systems (KBSs) that operate in complex domains must "reason" from information that is to some degree uncertain, imprecise, and occasionally inaccurate, called "evidential" information [25]. Furthermore, each body of information is usually generically distinct and is typically obtained from a variety of disparate sources, commonly called knowledge sources (KSs). The evidential information that KSs provide is derived, in part, from imperfect perceptions of their environment. And as such, can be viewed as partial evidence for or against the occurrence of semantically meaningful events in some domain of interest. Given this reality, the degree to which a KBS successfully deals with real world problems depends, in part, on the technology it employs to reason from evidential information.

In this paper, we are concerned with the integration and evaluation of a technology that KBSs might use to complete two fundamental tasks. One task is to reason from evidential information in order to interpret (i.e., understand) the perceptions of its KSs. The second is to decide how to allocate its limited resources in order to successfully complete the previous task. That is, we must anticipate that the complexity of the real world prohibits a KBS from understanding its perceptions in one fell swoop. Rather, "control-related" information must be obtained in order to help make decisions about the type, nature, quality, and quantity of the information that is required to interpret the perceptions of KSs. In the work reported here, the control-related information that a KBS must reason from is provided by control knowledge sources (CKSs). Similar to KSs, the information that CKSs provide is derived, in part, from their perceptions of the state of the system and or the environment. As a consequence, such control related information is also evidential in nature. Thus, KBSs will be more successful at understanding their perceptions to the degree they employ technologies that are better suited than current techniques for reasoning from limited evidential information.

Recent research indicates that Dempster's rule for combining beliefs, and Shafer's theory of belief functions shows promise as a more viable alternative than some popular Bayesian based techniques for addressing these problems [7], [34], [26], [14], [8], [27], [15], [38], [1], [16], [17], [41], [42], [43], [44]. In addition, the work of Dempster and Shafer is the basis of a developing concept called "evidential reasoning" (ER) [25]. This concept, which we shall discuss later, is the foundation of our framework for addressing the above interpretation and control problems in our domain of interest.

It is clear that the problem of reasoning from evidential information is common to many KBSs that operate in complex domains. However, we have chosen the task domain of high-level computer-based image interpretation as the context within which to discuss and present some of our work. Within this context, we shall discuss research on the application of the DS and ER technologies to a KBS that is designed to interpret two-dimensional monocular color images of outdoor natural scenes.

We begin the discussion by stating a major objective of general purpose high-level knowledge-based image interpretation systems. Next we briefly describe the "origins" of the evidential information that an image interpretation system must reason from in order to complete its tasks. Then we describe some of the difficulties with using probabilistic-based approaches for reasoning from evidential information. Following this discussion, we shall introduce Shafer's theory of belief functions, Dempster's rule, and contrast it with some aspects of probability theory. Next, we shall acquaint the reader with Lowrance's and Garvey's concept of evidential reasoning (ER) [25]. And after introducing this concept, we shall describe our high-level knowledge-based image interpretation system that was built to employ and explore some aspects of both the Dempster-Shafer (DS) theory and ER technology. Finally, results of interpretation experiments will be presented followed by a discussion of related and future work in this area.

We mention, here, that our emphasis throughout this paper shall be on the underlying technology a KBS might use to reason from evidential information. For examples and a

more extensive review of computer vision systems see, for instance, Nagao and Matsuyama [28], Binford [3], Havens and Mackworth [20], Selfridge [33], Brooks [4] Sloan [37], Hanson and Riseman [19], Levine [23], and Levine and Shaheen [24].

2 IMAGE INTERPRETATION OBJECTIVES

An example of a typical complex outdoor natural scene that a general purpose knowledge-based image interpretation system might be expected to understand is shown in Figure 1. An objective of such systems is to identify semantically meaningful visual entities in a digitized and segmented image of some scene. That is, to correctly assign semantically meaningful labels (e.g., house, tree, grass, and so on) to regions in an image – see [29], [30]. A computer-based image interpretation system can be viewed as having two major components, a "low-level" component and a "high-level" component [19], [31]. In many respects, the low-level portion of the system is designed to mimic the early stages of visual image processing in human-like systems. In these early stages, it is believed that scenes are partitioned, to some extent, into regions that are homogeneous with respect to some set of perceivable features (i.e., feature vector) in the scene [6], [40], [39]. To this extent, most low-level general purpose computer vision systems are designed to perform the same task. An example of a partitioning (i.e., segmentation) of Figure 1 into homogeneous regions is shown in Figure 2. The knowledge-based computer vision system we shall describe in this paper is not currently concerned with resegmenting portions of an image. Rather, its task is to correctly label as many regions as possible in a given segmentation.

It is clear that no segmentation is perfect. There will be regions that overlap semantically distinct visual entities. Or there might be regions that are over segmented – i.e., multiple regions that partition a single semantic entity. These anomalies are due, in part, to several unavoidable realities of the visual domain. Imaging machinery will simultaneously lose meaningful information and introduce bogus information – e.g., noise and or distortion. Thus, the data from which a segmentation must be produced is an imperfect abstraction of the scene a system is expected to understand. Second, semantic

information about objects in a scene cannot be contained entirely in the image data. And as a consequence, some regions will partition a single visual entity. And still other regions may enclose multiple semantically distinct visual entities.

In the KBS we shall be describing, KSs extract a variety of image feature information from a subset of regions in a segmented image – e.g., spectral, texture, shape, and spatial attributes of regions. Based on their perceptions, KSs form opinions about the presence and or absence of features they are capable of observing. What logically follows is that beliefs that are based on these opinions will be imperfect. And at best such opinions can be viewed as only *evidence* to suggest the presence or absence of semantically meaningful entities in a particular scene of interest.

Given this reality, how might a system represent the evidential information it obtains from KSs? And how might a KBS reason from this evidence more effectively than current approaches permit? Let us begin to answer this question by briefly reviewing current approaches.

2.1 CURRENT APPROACHES TO REASONING

Some of the problems with reasoning from evidential information are common to many domains other than computer-based image interpretation. Many of the currently popular approaches for addressing them are probabilistic in nature. That is, probabilities are used to represent belief in propositions and Bayes rule or an ad hoc variant thereof is typically used to update a system's belief in propositions based on new bodies of evidence. See, for instance, the work on systems like VISIONS [19], Prospector [10], MYCIN [36], and Gorry's computer-aided medical diagnosis system [18]. The problems with probabilistic based approaches are well known and continue to be discussed in the literature [35], [26], [21]. However, let us briefly state a few of the problems that motivated our exploration into alternative theories.

One concern with probabilistic based models is with their voracious appetite for data. Where H and e represent some hypothesis and body of evidence, respectively, in order to use the inversion formula of Bayes rule we must have some prior belief $P(H)$, $P(e)$, and likelihood $P(e|H)$. It is well known that the complexity of real world domains makes it difficult, at best, to obtain or reliably estimate such beliefs and likelihoods. Some have countered by saying that a complete probability specification is not required [32]. Rather, one need only estimate the *odds* or *likelihood-ratio* – see [32]. Our concerns with this view is that a large number of likelihood-ratios still must be provided, and that the problems of not having a uniform represention of ignorance and being able to distinguish disbelief from no belief also remain.

In a probabilistic approach, one typically represents ignorance in a set, say Θ, of mutually disjoint propositions by the following probability distribution P_0:

$$\underset{\theta\in\Theta}{P_0(\theta)} = \frac{1}{|\Theta|}. \tag{1}$$

If it becomes necessary to change Θ to Θ', where $|\Theta| \neq |\Theta'|$, then our numerical representation of ignorance must also change. Such a change might have been induced by the acquisition of additional evidence. If this is the case, by what theory does one reconcile or interpret the disparity in the representation of ignorance in Θ and Θ'? It is important that a system capable of dealing with such disparities, particularly when it happens to be equally ignorant about some $\theta \in \Theta$, and the same $\theta \in \Theta'$, but $\underset{\theta\in\Theta}{P_0(\theta)} \neq \underset{\theta\in\Theta'}{P_0(\theta)}$.

The additivity constraint

$$P(A) + P(\neg A) = 1 \tag{2}$$

imposes some undesirable restrictions on a system's ability to distinguish *disbelief* from *no* belief in the truthfulness or falseness of a proposition. If our belief in A happens to be, for instance $P(A) = x$, then we are forced to believe to a degree of $1-P(A) = P(\neg A) = 1-x$ that $\neg A$ is true. It might be the case, however, that we have no evidence to indicate $\neg A$

is true or A is false to any degree. Thus, we must adopt beliefs for which we have no evidence to support. And as a consequence, we cannot distinguish, in a nice single formal representation, disbelief from no belief.

To summarize our concerns, a pure probabilistic based approach to reasoning in complex domains is overly restrictive. It is difficult to specify a complete probability distribution over the propositions of interest due to the enormous number of micro events that must be taken into account. And as a consequence, pure probabilistic approaches are typically compromised by making ad hoc modifications to Bayes' rule in order to deal with these restrictions – see for instance [9], [10]. A formal and uniform representation of ignorance remains unavailable. And we cannot distinguish disbelief from no belief. These are a few of the unwarranted constraints that have motivated us to seek alternative approaches to these problems. The results of our efforts have led us to investigate some of the work of Arthur Dempster and Glenn Shafer – commonly called the Dempster-Shafer (DS) theory [7], [34].

3 THE DEMPSTER-SHAFER THEORY

We can view the problem of reasoning in complex domains as one of trying to answer a particular question of interest. For instance, in the computer vision domain a system might be asked to solve the problem of identifying an object in some region of interest. A typical question might be which of the following disjoint propositions, say θ_1 and θ_2 is true: *The region is a house* (θ_1) or *The region is a barn* (θ_2)? Or in other words, which label hypothesis, *house* or *barn*, should be associated with the current region of interest? A system must answer this question, in part, by obtaining and pooling the appropriate beliefs that would allow it to discern a house from a barn. Then the problem is solved – i.e., the question is answered – to the extent a system is capable of successfully obtaining and pooling such beliefs.

In Shafer's theory, the degree of belief, Bel, that one should accord a proposition is represented as a number between zero and one. Suppose Θ is a finite set, and we denote the set of all subsets of Θ by $\wp(\Theta)$. Then if Bel satisfies the following conditions:

(1) $Bel(\emptyset) = 0$.

(2) $Bel(\Theta) = 1$.

(3) For every positive integer n and every collection $A_1, \ldots, A_n$ of subsets of Θ,

$$Bel(A_1 \cup \ldots \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + - \ldots + (-1)^{n+1} Bel(A_1 \cap \ldots \cap A_n), \quad (3)$$

then Bel is a *belief function* over Θ. Within the context of Shafer's theory and with respect to this simple example, $\Theta = \{\theta_1, \theta_2\}$ and is called a *frame of discernment.* It is clear that what constitutes Θ and how it is used is crucial to the success of problem solving systems. Therefore, let us provide more background about a frame of discernment before returning to our discussion of belief functions.

3.1 FRAME OF DISCERNMENT

Suppose we are presented with a question and a finite set, Θ, consisting of possible answers to the question, only one of which is the correct one. Then for each $\theta \in \Theta$ the propositions of interest are precisely those of the form "*The correct answer is* θ_1", "*The correct answer is* θ_2", ... , "*The correct answer is* θ_n", and so on for $n = |\Theta|$. Simply stated, a set is called a *frame of discernment* when its elements are interpreted as possible answers to a particular question, and we know that exactly one of these answers is correct. And what logically follows from this statement is that the set of all propositions of interest are in a one-to-one correspondence with the set of subsets of Θ, – i.e., $\wp(\Theta)$.

A frame of discernment is epistemic in nature. Its meaning and justification for existing lies in the knowledge and evidence that is brought to bear in order to *discern* the correct answer – i.e., which singleton proposition in Θ is true. One will be able to identify the correct answer to a question only to the degree that a frame adequately captures the relevant interaction of such knowledge.

Consider the large amount of information that is typically required to answer any particular question in the real world. For example, suppose a robot were trying to answer the question what is the object currently in its field of view. A set of possible answers to this question might be *"A House"*, *"A Barn"*, *"A Tree"*, and so on. Examples of knowledge that might be brought to bear on this question are: the shape of houses, barns, and trees; the texture of each object; the spectral attributes of houses, barns, and trees; and perhaps the spatial relationships between these objects. Each example just given is a generically distinct type of knowledge and, by itself, can be viewed as a relatively "small world" of knowledge compared to the total amount that is usually needed to answer this and more complex questions.

The propositions contained in each distinct body of knowledge might relate quite differently to subsets of the possible answers. For instance, with respect to shape, if the proposition *The region contains many lines meeting at obtuse angles to one another* were true, then we would want to admit that it is possible *"A House"* or *"A Barn"* is the correct answer and that *"A Tree"* is not. Similarly with respect to texture, houses and barns might be relatively less textured than trees. Then if the proposition *The region is relatively smoothly textured* is true then we would want to further admit that *"A House"* or *"A Barn"* is possibly a correct answer and *"A Tree"* is not.

In this example, each proposition in each distinct small world can be viewed as a "feature proposition" (e.g., *"The region is relatively smoothly textured"*) that might help to discern which answer is possibly correct. The set of all feature propositions in a small world constitutes a "feature space" (e.g., the texture feature space). Each feature

space can be thought of as containing at least one proposition that is associated with some observable and quantifiable aspect, called a "feature value", of the related chunk of knowledge (e.g., the average number of obtuse angles formed by straight lines in a region). The set of all feature spaces of potential interest constitutes an "environment." And in general, this includes any aspect of a domain or world about which information may be obtained in order to help decide which answer is correct. With this partial background we can present a more formal description of a frame of discernment.

3.1.1 A formal view. Let the set of mutually exclusive and exhaustive possible interpretations of an image be represented by the set Θ_Q, where

$$\Theta_Q = \{\theta_1, \theta_2, \ldots, \theta_n\}. \tag{4}$$

We can associate with each θ_i, $1 \leq i \leq n$, a proposition that represents an interpretation – e.g., θ_1 might be associated with the proposition *The image is a house scene*, θ_2 might be associated with the proposition *The image is a tree scene*, or θ_3 might be associated with the proposition *The image is a house and tree scene*, and so on.

Let $F_1, F_2, \ldots, F_m$ correspond to the feature spaces of interest – e.g., F_1 might correspond to the spectral features of objects, F_2 might correspond to the texture features of objects, and so on. Associated with each F_i, for $1 \leq i \leq m$, is a set $\mathcal{F}_i$ of possible feature values of F_i,

$$\mathcal{F}_i = \{f_i^k \mid f_i^k \text{ is a possible feature value of } F_i, \text{ for } 1 \leq k \leq |\mathcal{F}_i|\}. \tag{5}$$

For example, if $\mathcal{F}_1$ is the set of feature values for the texture feature space F_1, then f_1^1 may correspond to a relatively smooth texture value, as might be characteristic of objects such as sky or paved road. Similarly, f_1^2 may correspond to a relatively rough texture value, as might be characteristic of objects such as tree crowns, or grass. Like each θ_i, we can also associate with every f_i^k a proposition that describes a possible outcome

as a result of performing a perceptual operation–e.g., f_i^1 might be associated with the proposition *The image contains relatively rough textured regions*, f_1^2 might be associated with the proposition *The image contains relatively smooth textured regions*, or f_1^3 might be associated with the proposition *The image contains both relatively rough and smooth textured regions*, and so on.

For each $f_i^k \in \mathcal{F}_i$ it is possible to identify a subset of Θ_Q that *possibly* contains the correct interpretation when f_i^k is observed. For instance, let Θ_Q be defined as follows:

$$\Theta_Q = \{ \text{ tree crown, sky, grass, paved road} \}. \tag{6}$$

Let f_1^1 and f_1^2 correspond to the texture feature values just discussed. If observations by a texture KS indicate that the proposition *The image contains relatively smooth textured regions* (i.e., f_1^2) is true, then it is possible that the region of interest should be labeled sky, or paved road and should *not* be labeled tree crown, or grass. Conversely, if observations by a texture KS indicate that the proposition *The image contains relatively rough textured regions* (i.e., f_1^1) is true, then it is *possible* the measured region should be labeled tree crown, or grass and should *not* be labeled sky, or paved road.

Given that we can identify a subset of Θ_Q that is possibly the correct answer for a question of interest when a feature $f_i^k \in \mathcal{F}_i$ is observed, The set Θ_Q can be generated by a characteristic set function that is defined over the space of feature propositions of interest – i.e., the $f_i^k s \in \mathcal{F}_i$ s. That is, we can define a distinct function χ_i over each $\mathcal{F}_i$ to be:

$$\chi_i : \mathcal{F}_i \mapsto \wp(\Theta_Q), \tag{7}$$

such that

$$\bigcup_{f_i^k \in \mathcal{F}_i} \chi_i(f_i^k) = \Theta_Q. \tag{8}$$

χ_i is called the *characteristic set function* of $\mathcal{F}_i$, and $\chi(f_i^k)$ is called the *characteristic set* of f_i^k. An example characteristic function for our example above might be:

1) $\chi(\textit{The region is relatively smoothly textured}) = \{\textit{sky, paved road}\}$;

2) $\chi(\textit{The region is relatively rough}) = \{\textit{tree crown, grass}\}$.

A frame of discernment, then, can be defined in terms of a set Θ_Q and a collection of feature spaces and their characteristic set functions.

A frame is said to be *internally complete* if every element of Θ_Q can be realized as an intersection of characteristic sets. Note that if perfect information is available (i.e., every feature proposition is either true or false) and Θ_Q captures the relevant interaction of our knowledge and available evidence, then drawing inferences over Θ_Q amounts to computing set intersections. For instance, with respect to the question which proposition, $\theta \in \Theta_Q$, is true:

1) if f_1^k is true then the correct answer (i.e., the proposition $\theta \in \Theta_Q$ that is possibly true) lies in $\chi_1(f_1^k)$;

2) if $f_2^{k'}$ is true then the correct answer lies in $\chi_2(f_2^{k'})$.

And as a consequence, the combined correct answer is contained in

$$\chi_1(f_1^k) \cap \chi_2(f_2^{k'}). \tag{9}$$

3.2 CONVEYING OPINIONS

In this scheme a KS might convey its opinion about the degree to which it believes feature propositions are true or false through the following "mass function" M:

$$M : \wp(\Theta_Q) \mapsto [0,1], \quad \text{where} \quad M(\emptyset) = 0, \tag{10}$$

and

$$\sum_{A \subseteq \Theta_Q} M(A) = 1. \tag{11}$$

We note here that a Bayesian probability distribution m:

$$m : \Theta_Q \mapsto [0,1], \quad \text{where} \sum_{\theta \in \Theta_Q} m(\theta) = 1, \tag{12}$$

is just a special case of a mass function. And that both M and m satisfy the conditions of equation 3, and are belief functions over Θ_Q. The implication of this statement is that if a complete probability specification is available, then the DS theory is capable of integrating this information with other bodies of evidence.

Each body of evidence induces an interval, called an "evidential interval", within which belief about a proposition must lie. An evidential interval is a subinterval of the real interval [0,1]. The lower and upper bounds of the evidential interval shall be called the support (Spt) and plausibility (Pls), respectively. The Spt represents the total mass that tends to support a proposition:

$$Spt(B) = \sum_{p \subseteq B} M(p). \tag{13}$$

The Pls represents the degree to which the mass fails to refute the proposition.

$$Pls(B) = 1 - Spt(\neg B) = 1 - \sum_{p \subseteq \neg B} M(p). \tag{14}$$

The degree to which the mass refutes a proposition is called the dubiety (Dbt).

$$Dbt(B) \quad = \quad Spt(\neg B) \quad = \quad 1 - Pls(B). \tag{15}$$

The degree to which no mass tends to support a proposition or its negation is called ignorance (Igr).

$$Igr(B) \quad = \quad Pls(B) - Spt(B). \tag{16}$$

The interpretations of some evidential intervals are summarized below:

Completely true proposition $[1,1]$;

Completely false proposition $[0,0]$;

Completely ignorant about the proposition $[0,1]$;

Tends to support the proposition $[Spt,1]$, $0 < Spt < 1$;

Tends to refute the proposition $[0,Pls]$, $0 < Pls < 1$;

Tends to both support and refute the proposition $[Spt,Pls]$, $0 < Spt \leq Pls < 1$.

Note that with a mass function a KS is able to express its beliefs at any desired precision or certainty – i.e., a source can express beliefs by attributing any amount of mass to any proposition it desires. In addition, an evidential interval allows one to distinguish disbelief from no belief, unlike a pure point probabilistic representation. And finally, an evidential interval provides a single formal and uniform representation of ignorance – i.e., the interval [0, 1] always represents total ignorance across model variations.

3.3 COMBINING BELIEF (MASS) FUNCTIONS

Given the complexity of the real world, it is unlikely that a single source of information will be capable of providing independent beliefs (i.e., opinions) about the truthfulness or falseness of feature propositions. Rather, a more pragmatic approach is to have multiple distinct KSs express opinions about their perceptions by attributing a portion of their unit mass to feature propositions. In this approach, we need to be able to form a consensus opinion by combining multiple mass functions. In the theory of belief functions, the tool for carrying out this pooling process is Dempster's rule [7].

3.3.1 Dempster's rule. Dempster's rule tells us how to take two mass distributions M_1, M_2 and produce a third mass distribution M_3 that represents a consensus opinion of two distinct sources, and is defined to be:

$$\text{For all} \quad B_1, B_2, B_3 \subseteq \Theta_Q, \quad M_3(B_3) = (1-K)^{-1} \sum_{B_1 \cap B_2 = B_3} M_1(B_1) M_2(B_2),$$
$$\text{for} \quad K = \sum_{B_1 \cap B_2 = \emptyset} M_1(B_1) M_2(B_2) \leq 1, \tag{17}$$

where K is the total amount of conflict between M_1 and M_2, and $(1-K)^{-1}$ is a renormalization factor.

Using Dempster's rule to combine mass distributions accomplishes three functions. The first is to obtain a consensus about what answer each source believes is possibly correct. If both opinions are completely consistent, there is at least one answer that both sources agree is correct, and it can be said they are expressing totally compatible opinions – i.e., $K = 0$. Conversely, if beliefs are not completely consistent, then their opinions are not totally compatible and there is at least one answer the sources disagree is appropriate– i.e., $0 < K \leq 1$. In general, to the degree that sources are certain, precise, and accurate with respect to their observations, their opinions about which answers are correct will

be compatible. Dempster's rule determines simultaneously if there is any $\theta \in \Theta_Q$ that multiple sources agree is true and provides a measure of compatibility among the opinions they provide.

The second function of Dempster's rule is to correct for minor errors. The assumption here is that there is only a negligible likelihood that distinct sources might introduce the same type of error into their opinions simultaneously – i.e., that they are stochastically independent. Therefore, such errors can be overcome by a sufficient amount of redundant and generally correct beliefs. If a subset of sources make gross errors, such bad information should be discounted, when detected.

The third function of Dempster's rule is to compute the minimum degree of support that should be attributed to compatible opinions, if such an opinion exists. In a sense, the multiplicative nature of equation 17 computes the minimum commitment of support one should attribute to compatible opinions that were provided from independent sources. But the requirement that sources be independent is crucial to the applicibality of the rule and is discussed in the following section.

3.3.2 Independence. Consider the following excerpt from the paper that describes the independence requirements of Dempster's rule [7].

"... Opinions of different people based on overlapping experience could not be regarded as independent sources. Different measurements by different observers on different equipment would often be regarded as independent, but so would different measurements by one observer on one piece of equipment: here the question concerns independence of errors."

Our reason for presenting this excerpt is to emphasize that the independence constraint that must be satisfied before Dempster's rule is potentially applicable is with respect to the *errors* multiple sources might make. And despite that fact that this point has been made in the mathematical literature, it has escaped recognition by a significant portion

of the artificial intelligence community that relies, to some degree, on the DS calculus for reasoning from evidential information.

This notion of independence is quite different from that which typically comes to mind during discussions of classical probabilistic models for pooling information. The classical definition of stochastic independence for n events, E, is defined as:

$$P(\bigcap_{j=1}^{n} E_j) = \prod_{j=1}^{n} P(E_j). \tag{18}$$

But we must be careful when we try to interpret this equation in the context of "... here the question concerns independence of errors.", [7]. This is so because both the Bayesian theory and belief function theory treat chance in different ways, and as a consequence their concepts of independence are slightly different.

Consider two propositions $A, B \subset \Theta_Q$ that, for the moment, happen to be false with respect to a particular frame and bodies of evidence. Now let E_k be the event that KS_k attributes a non zero amount of mass in support of A. And let E_i be the event that KS_i attributes a non zero amount of mass in support of B, where $A \cap B \neq \emptyset$ and $1 \leq i \neq k \leq n$. That is, for a particular frame both KSs have simultaneously errored in their assessment of some body of evidence. Then $P(E_k \cap E_i)$ is interpreted as the probability or chance that both KS_k and KS_i will simultaneously express opinions that are compatible *and* erroneous.

If a frame of discernment has taken into account all significant dependencies then the left hand side of equation 18 will, as a consequence, be zero. Also notice that this does not mean that all sources must be error free in order for this equality relation to hold. It is only necessary that at least one $P(E_j) = 0$.

Under less than ideal conditions we need to augment the meaning of an event E_j slightly. The reason is because noise is random in nature, and as a consequence we must

anticipate that equation 18 will not always be zero. Thus a more realistic concern is with the probability or chance distinct sources will simultaneously introduce errors above some noise threshold, say t, into their opinions. Now we can interpret an event E_k to mean that KS_k will attribute an amount of mass (i.e., support) $m_1(A) > v_k$, given that a source KS_i will attribute an amount of mass $m_2(B) \leq v_i$ where $0 \leq t \leq v_i, v_k \leq 1$ and $A \cap B \neq \emptyset$. Then $P(E_k)$ can be interpreted as the *a priori* probability or chance that KS_k will introduce errors larger than v_k into its assessment of some body of evidence. That is, we are only concerned with the chance that independent sources will simultaneously introduce errors above some "noise" level in their opinions. We have shown how one can determine the maximum amount, v_k for instance, of mass that KS_k can attribute to a false proposition, say A, and keep the total amount of $Spt(A \cap B)$ below some level s, given that KS_i attributes an amount of mass in support of, say B, below v_i [44].

With respect to both the Bayesian and DS technologies, as well as many others, one tries to make the relevant dependencies of the current problem effectively independent within the context and constraints of each theory. Accomplishing this makes the machinery of each model potentially appropriate to use. With this background we are now ready to acquaint the reader with the concept of evidential reasoning.

4 EVIDENTIAL REASONING

The concept of evidential reasoning (ER) was introduced by Lowrance and Garvey [25]. This evolving technology starts from the position that the acquisition of information by KBSs involves making imperfect perceptions of the environment. A KBS "understands" its world by perceiving it through a set of KSs. And because a system's perceptual machinery is not flawless, it follows that the information the KSs provide will be to some degree uncertain, imprecise, and occasionally inaccurate – evidential in nature. This concept currently relies on the DS formalism as its model for representing and pooling KSs' beliefs that are based on their environmental perceptions. Thus the DS formalism is fundamental to ER-based models that KBSs might use to reason in their task domain.

There are two distinct reasoning processes that must be completed in this concept. One is to take a single body of evidence and propagate its effect from those propositions the evidence bears directly upon to those it indirectly bears upon. This allows inferences to be drawn about those propositions not directly affected by the evidence. This process is typically carried out by what is commonly called an inference engine. The other process, one that pools multiple bodies of evidence into a single body of evidence that represents a consensus opinion, is Dempster's rule which we have already described.

We can summarize these processes in terms of a KBS's two computational requirements, which for $B, C \subseteq \Theta_Q$ are:

1.) **Combination of multiple** M**'s:**

   a) Apply Dempster's rule to M_1 and M_2 to produce a consensus opinion that is reflected in $M_3 = M_1 \oplus M_2$.

   b) If $B \cap C \neq \emptyset$, then add $M_1(B) * M_2(C)$ to current $M_3(B \cap C)$.

   c) If $B \cap C = \emptyset$, then add to current k.

2.) **Extrapolation**: Taking the result of Dempster's rule (i.e., M_3) and computing the Spt and Pls of the remaining dependent propositions.

   a) If $B \subseteq C$ then add $M(B)$ to current $Spt(C)$.

   b) If $B \subseteq \neg C$ then add $M(B)$ to current $1 - Pls(B)$.

With this fairly extensive discussion of Dempster's rule, Shafer's theory, and the concept of evidential reasoning, we can now introduce our evidential-based high-level computer vision system (EHCVIS). After which we shall show how we have used both the DS and ER technologies to reason in our task domain.

5 EHCVIS

A flow diagram of EHCVIS and a generalized illustration of its architecture is shown in Figure 3. We do not claim that the architecture of our system is ideal for a high-level computer vision system – see, for instance, Levine for a discussion of a general design for computer vision systems [24]. Rather, the design we have chosen is one of perhaps many that might be adequate for interpreting images and exploring various aspects of the DS theory.

As indicated by the figure, EHCVIS can be described in four phases. The task of the first phase is to use the specifications of goals to help complete two subtasks. Examples of goals the system might try to reach are finding a house, locating the ground plane, or obtaining additional information to help resolve some ambiguity the system might have about the identity of objects in a region of interest. The first subtask is to use goal specifications to generate a set of alternative actions the system might pursue in order to reach that goal. The second subtask is to use goal specifications to select control strategies that will be used to help decide which alternative action is more appropriate to pursue.

The second phase can be summarized in several steps: (1) with the alternative actions and control strategies that were selected in the previous phase, dynamically build the control knowledge (i.e., Θ_A) that will be brought to bear on the problem of deciding which alternative to pursue; (2) implement these control strategies, in part, by obtaining control related information from independent control knowledge sources (CKSs); and (3) pool these beliefs using Dempster's rule and then use an inference engine to take the result of Dempster's rule and infer which action is the best to pursue. Note that in the third step of the second phase (see Figure 3) beliefs are pooled and inferences are drawn over the frame denoted by Θ_A. This is to indicate that the DS technology is used by our system to reason about its actions.

In the third phase, our system takes the action suggested by the second phase. A typical action might be to task a subset of available KSs to make some observation about a particular subset of regions in an image, then express some beliefs about their perceptions. After the KSs have done this, Dempster's rule is used to pool their beliefs and then inferences are drawn over Θ_Q to infer which propositions (i.e., label hypotheses) in Θ_Q should be associated with the region under examination.

In the last phase, the results of the inferences drawn over Θ_Q are evaluated. Based on this evaluation the system might decide that a new goal should be satisfied and return to the goal generation phase. Or that the interpretation process should be terminated. Or that the system should "instantiate" (i.e., record in a dynamic representation called short term memory, STM) its belief that a subset of the label hypotheses in Θ_Q should be associated with a subset of the regions in an image, and then set new goals to be satisfied. Let us briefly discuss each phase in more detail.

5.1 PHASE ONE:

5.1.1 Goals. EHCVIS begins the interpretation process when a goal is placed on a goal stack. Every goal contains three parts: (1) a symbol that indicates a goal-name or goal-identification. For example, `verify-kss`, and `reduce-ignorance-about-a-hypothesis` are examples of symbols that indicate the goal of verifying the preception of KSs, and indicate the goal of reducing the system's ignorance about the truthfulness or falseness of a label hypothesis, (2) a specification of a set of KS selection constraints; and (3) a specification of a set of region selection constraints. The KS selection constraints specify attributes of KSs that must be satisfied before the system will consider them as potential sources of information. Similarly, the region selection constraints specify characteristics that regions must possess before the system will consider obtaining information about them. For instance, when the system begins the interpretation process it is totally ignorant about the identity of objects that are depicted in an image. One "`start-up`"

goal might be to obtain preliminary information about a subset of the regions in the image. It might be desirable to satisfy this goal by tasking the most reliable KSs to obtain information about "unusual" regions – i.e., regions exhibiting features that might cause, say a human, to foveate to upon initial examination of an image. An example of how such a goal and its selection constraints might be specified is shown in Figure 4. The symbol `start-up`, in Figure 4 indicates that the system should try and reach the goal of obtaining preliminary information about some region in an image. The symbol `rel` in the list `(rel 0.7 1)` of the KS selection constraint portion of the `start-up` goal indicates that this constraint pertains to the reliability of KSs. And the interval `(... 0.7 1)` indicates the range within which the reliability of a KS must lie before it is considered a potential source of information. How the reliability of KSs might be determined is not of interest at the moment and is discussed in more detail in [44].

Upon initiation of the interpretation task it might be desirable to focus attention on unusual regions – e.g., regions that are relatively large, relatively bright or dark, or at some extreme location in an image. Suppose our system is initially interested in relatively large regions at a relatively high location in an image. The following region selection constraint might be used to specify this interest:

```
(conj ((loc-above x y)
       (size min-size max-size) ...)).
```

Where `(conj ((loc-above ...) (size ...) ...))` means a region becomes a potential candidate for examination if its location is above some minimum `x y` position in the image and its `size` is within the range `(... min-size max-size)`. The shaded region of the image in Figure 4 exemplifies the result of applying a similar constraint to an image that has been interpreted by our system.

5.1.2 **Generating Alternatives.** The output from the KS and region selection process is a list of KSs to possibly task and a set of regions in the image these KSs might be asked to obtain information about. It might be necessary or desirable to task multiple KSs from the list of candidate KSs – e.g., $\mathrm{KS}_1 \wedge \mathrm{KS}_2$ – on a collection of regions – e.g., $R_5 \cup R_{50}$. If we let k and r represent the list of selected KSs and regions respectively, the system generates the sets:

$$K \subseteq \wp(k) \qquad \text{and} \qquad R \subseteq \wp(r). \tag{19}$$

A typical $\kappa \in K$, might be the set $\{KS_1,\ KS_2\}$, and should be interpreted to mean $KS_1 \wedge KS_2$. Similarly, a typical $\rho \in R$, might be the set $\{R_3,\ R_{14},\ R_6\}$, and should be interpreted to mean $R_3 \cup R_{14} \cup R_6$. If k and or r are large, the pragmatics of generating K and R could be prohibitive. However, in practice we might know a priori that some KSs cannot be simultaneously tasked, thus eliminating some possibilities. Sometimes the KS or region selection constraints might keep the size of k and r relatively small. Other times, however, k and r can remain relatively large. When this is the case, EHCVIS randomly choose a manageable subset of k and r to work with. The size of the subset chosen is a function of the available computational resources. And we have pointed out that systems must perform a similar operation when the amount of data becomes overwhelming and the information to help prune the choices is not available [43].

Once K and R have been generated the frame of discernment, Θ_A, from which EHCVIS must choose an alternative is defined to be:

$$\Theta_A \subseteq \{invoke\text{–}\} \times K \times R. \tag{20}$$

For example, consider the following propositions in Θ_A:

$$\begin{aligned} \Theta_A = \{&invoke - KS_1 \wedge KS_2 - R_3, \\ &invoke - KS_5 \wedge KS_1 \wedge KS_2 - R_3 \cup R_{14},\ \ldots\}. \end{aligned} \tag{21}$$

We interpret an alternative, $\theta \in \Theta_A$, of the form $invoke - KS_1 \wedge \ldots \wedge KS_n - R_1 \cup \ldots \cup R_k$ to mean task KS_1 and KS_2 and ... and KS_n to simultaneously obtain information from the region formed by $R_1 \cup R_2 \cup \ldots \cup R_k$ – i.e., the region formed by the union of R_1 and R_2 and, ..., and R_k. But given a set of alternatives, what control strategies might a system employ to help decide which alternative to pursue?

5.1.3 Control strategies. EHCVIS has eleven "primitive" control strategies which can be used to help decide which alternative action to pursue. One primitive control strategy is to obtain information in support of or against hypotheses for which the system is most ignorant about. This strategy might be used to reduce the system's ignorance in the truthfulness or falseness of a label-hypothesis. A second strategy is to obtain information that will help to reduce the system's ambiguity about the truthfulness or falseness of a subset of label hypotheses in Θ_Q. More complex control strategies are formed by "merging" two or more primitive control strategies. The details of this merging process will be discussed shortly.

Currently, EHCVIS uses a simple "table-driven" scheme to decide which control strategies should be used. For each goal the system is expected to reach, there is an entry in a table that lists a subset of the available primitive control strategies that should be used to help reason about what action to pursue. For example, with respect to the `start-up` goal, EHCVIS's table currently indicates that two primitive control strategies should be simultaneously used: the strategy of invoking the most reliable KS, and the strategy of obtaining information about hypotheses the system is most ignorant about.

A more complex strategy is specified by enumerating two or more primitive control strategies in this table. Now let describe how they are effectively implemented within our control framework.

5.2 PHASE TWO:

5.2.1 **Building Θ_A.** Each primitive control strategy can be viewed as a "control feature space." Each control strategy is associated with a small world of control knowledge that might be brought to bear on the question of which action to take. Suppose we let F_1 represent the control feature space that is related to the reliability of KSs. Then within the context of the DS theory, we must enumerate the set, $\mathcal{F}_1$, of control feature propositions that are associated with F_1, and then define the characteristic function:

$$\chi_1 : \mathcal{F}_1 \mapsto \wp(\Theta_A). \tag{22}$$

EHCVIS enumerates $\mathcal{F}_i$s and constructs χ_is dynamically because, unlike Θ_Q, the set of alternative actions a system might pursue, as represented by Θ_A, typically cannot be know a priori.

EHCVIS dynamically builds Θ_A in the following manner. Suppose a_1, a_2, and a_3 correspond to the following alternatives in Θ_A:

$$\begin{aligned}\Theta_A = \{&invoke - KS_1 \wedge KS_2 - R_3\ (a_1),\\ &invoke - KS_5 \wedge KS_1 \wedge KS_2 - R_3 \cup R_{14}\ (a_2),\\ &invoke - KS_5 \wedge KS_{10} - R_{14}\ (a_3)\}.\end{aligned} \tag{23}$$

Then the following control feature propositions of $\mathcal{F}_1$ will be dynamically constructed:

f_1^1: *The most reliable KSs are* KS_1 *and* KS_2;

f_1^2: *The most reliable KSs are* KS_5 *and* KS_1 *and* KS_2;

f_1^3: *The most reliable KSs are* KS_5 *and* KS_{10}.

The reason these particular propositions have been enumerated is that prior to obtaining any information about the reliability of KSs, it is possible the KSs specified in each alternative might actually be the most reliable. Therefore each $f_1^k \in \mathcal{F}_1$, where $1 \leq$

$k \leq |\mathcal{F}_1|$, reflects this possibility, and it is the task of a CKS to dynamically measure this control-feature and then express its belief about which action, if taken, would result in tasking the most reliable KSs.

The function χ_1 can be defined, in words, to be as follows. For each $f_1^k \in \mathcal{F}_1$ extract the KSs the control-feature proposition claims is the most reliable – e.g., f_1^3 claims KS_5 and KS_{10} are the most reliable. Then include in the characteristic set of each f_1^k those actions that specify the same set of KSs. For example, $\chi_1(f_1^1) = \{a_1\}$, $\chi_1(f_1^2) = \{a_2\}$, and $\chi_1(f_1^3) = \{a_3\}$.

If there were a fourth alternative, say $a_4 \in \Theta_A$, that was defined to be $invoke - KS_1 \wedge KS_2 - R_3 \cup R_1$ then the characteristic set of f_1^1 would be $\chi_1(f_1^1) = \{a_1, a_4\}$.

If our reliability CKS believes that KS_5 and KS_1 and KS_2 are the most reliable, then it may express this belief by attributing a portion of its unit mass in support of f_1^2. Attributing more mass to f_1^2 the more it believed f_1^2 to be true, and less mass the less it believed f_1^2 was true. Or alternatively, attributing more mass to $\neg f_1^2$ the more it believed f_1^2 was not true.

Similarly, if the same CKS believes that KS_5 and KS_1 and KS_2 *or* KS_5 and KS_{10} are the most reliable, then it may convey this opinion by assigning a portion of its unit mass in support of the disjunction $f_1^2 \vee f_1^3$. Our reliability CKS may express total ignorance about which KSs it believes are the most reliable by assigning all of its unit mass to the disjunction $f_1^1 \vee f_1^2 \vee f_1^3$ – i.e., Θ_A.

Now consider a second primitive control strategy of obtaining information about regions that the system has the least information about. Again, associated with this strategy is a control feature space, say F_2, and a related set of control feature propositions $\mathcal{F}_2$. In this case, each $f_2^k \in \mathcal{F}_2$ would be of the form, for example, *The region the system knows the least about is* R_3, and so on for $1 \leq k \leq |\mathcal{F}_2|$. Next the system would define χ_2

in a similar manner as it defined χ_1 except that it would be with respect to the regions specified in each feature proposition and alternative. And our "region ignorance" CKS, like our reliability CKS, would be free to express any degree of support for or against any proposition or disjunction of propositions it desires. Now the frame Θ_A in this simple example is defined to be:

$$\Theta_A = \bigcup_{i=1}^{2} \chi_i(f_i^k). \tag{24}$$

And as a consequence of defining more than one characteristic function with respect to the same frame of discernment, a more complex control strategy has effectively been defined. That is, the strategy of tasking the most reliable KSs on the regions the system is most ignorant about.

If a frame of discernment is the mechanism by which complex control control strategies are defined. Dempster's rule is the machinery by which they are effectively implemented. If our reliability CKS believes the proposition $f_1^1 \vee f_1^2$ is true and our region ignorance CKS believes the proposition $f_2^2 \vee f_2^3$ is true then Dempster's rule determines if there exists an alternative action that both CKSs agree is appropriate to take. This is accomplished by intersecting the characteristic sets of the two propositions. In this instance, the action a_2 is the only alternative both CKSs agree the system should pursue. Thus, we have implemented the more complex strategy of tasking the most reliable KSs on regions the system is least knowledgeable about. In cases where a consensus opinion does not exist, Dempster's rule informs the system of this via its conflict measure k. When k becomes relatively large, a system must consider four possible causes: (1) one or more CKS expressed inaccurate opinions; (2) the frame is incomplete (i.e., some alternatives are missing from the model); (3) the goals are not satisfiable; or (4) a combination of the previous three.

5.2.2 How CKSs make measurements and convey beliefs. Consider two distinct types of control related information that CKSs might obtain:

1) **IGNORANCE-**(Igr): the total amount of evidence that neither supports nor refutes the truthfulness of a label hypothesis.

2) **AMBIGUITY-**(Amb): the total amount of evidence that *fails* to support or refute choosing a label hypothesis over its negation.

We defined the ignorance measure of a proposition, say $p \in \Theta_Q$, in equation 16. The ambiguity measure for a proposition, e.g., $p \in \Theta_Q$, is defined as:

$$Amb(p) = \begin{cases} Pls(p) - Spt(\neg p), & \text{for } Pls(p) \geq Spt(\neg p); \\ Pls(\neg p) - Spt(p), & \text{for } Pls(\neg p) \geq Spt(p); \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

In words, it is a measure of the amount of overlap of the evidential intervals of a proposition p and its negation, and thus represents the evidence that does not help to support or refute p.

Now that we know how some CKSs can measure the ignorance and ambiguity of propositions, how do they choose which control feature propositions, i.e., $f_i^k \in \mathcal{F}_i$, to support, and how much to support it? In EHCVIS, the specification of each KS contains a list of feature propositions, $f_i^k \subseteq \Theta_Q$ it can possibly express some opinion about. Just as some sensors in the real world can only perceive certain bandwidths of energy, some KSs can only observe certain features. As a consequence, each KS is capable of discerning only a subset of the label hypotheses of interest. A CKS uses the information in a KS's specification to help decide how much support to give to control feature propositions, $f_i^k \subseteq \Theta_A$. Let us provide an example of how this is accomplished.

Suppose our ambiguity resolving CKS has measured the ambiguity of, say ten, disjoint label hypotheses of interest and determined that the system is most ambiguous about

two of these label hypotheses, say p_4, $p_6 \in \Theta_Q$. Consider the following partial KS specification that might be used by our ambiguity CKS:

KS_1 : For the set of observable feature propositions $\mathcal{F}_1 = \{f_1^1, f_1^2\}$,

$$\chi_1(f_1^1) = \{p_2, p_3\},$$

$$\chi_1(f_1^2) = \{p_5\};$$

KS_5 : For the set of observable feature propositions $\mathcal{F}_5 = \{f_5^1\}$,

$$\chi_5(f_5^1) = \{p_5, p_6\}.$$

To reduce the ambiguity between p_4 and p_6 our system must obtain information from KSs that support either p_4 or p_6 but not both, or support either $\neg p_4$ or $\neg p_6$ but not both. We can see from the above partial KS specifications that KS_1 cannot provide such information. If KS_1 supports any subset of the feature propositions in $\mathcal{F}_1$ then both p_4 and p_6 become less plausible. Conversely, p_4 gains no support over p_6, or vise versa, if KS_1 refutes any subset of the feature propositions in $\mathcal{F}_1$. Unlike KS_1, if KS_5 gives support to any subset of its feature propositions then the amount of ambiguity between the two label hypotheses will be reduced. Thus, pursuing those actions that result in invoking KS_5 is more appropriate than pursuing those actions that invoke KS_1. The manner in which CKSs compute the degree to which any alternative should be supported is discussed in [44]. However, in a later section we shall present a simple example to illustrate the effect of the methods some CKSs use.

5.2.3 Decision criteria. As a consequence of CKSs expressing their opinions in terms of mass functions M, an evidential interval is induced over the alternatives in Θ_A. Selection of the appropriate action requires that these evidential intervals be evaluated. Although a complete classical utility theory for evaluating an interval representation of belief is not yet available, it is possible to choose actions on the basis of several simple criteria. For instance, the best action is obvious for those alternatives with nonoverlapping

intervals. For those choices with overlapping intervals, further evaluation is called for. There are many utility- vs. cost-based theories that might be used to select an action on the basis of beliefs that are constrained by an evidential interval. Although the details of how such theories might be employed are beyond the scope of this paper, we can describe the simple decision measure and criterion that EHCVIS uses.

This measure is motivated by the intuition that we should choose an alternative if the sum of the support for it minus the sum of the support for its competitors is greater than this same measure for the remaining alternatives. And the decision criterion used is to pursue the alternative that is indicated by the proposition having the largest value of the above measure. Since *Spt* and *Dbt* represent the sum of the support for and against a proposition, respectively, we can characterize this decision measure and criteria through the following equation:

$$\underset{a \in \Theta_A}{MAX}\ [Dec(a) = Spt(a) - Dbt(a)]. \tag{26}$$

For the case where this measure is the same for two or more alternatives, a random choice is made.

Unfortunately, an epistemological justification for this decision criteria cannot be offered at this time. Other people have suggested that just the plausibility, *Pls*, of an alternative is adequate [2]. However, our view is that further investigation might reveal that a combination of evidential measure might be more appropriate under different circumstances.

5.3 PHASE THREE:

5.3.1 Pursuing an alternative. A general purpose computer vision system must have a relatively large and sophisticated set of KSs in order to interpret images of complex scenes [19]. This is due, in part, to the need for a variety of information that typically cannot be provided by a single source. The process of building the necessary KSs remains an active area of research [45], [19], [11], [12], [22], [5]. And although there have been a number of significant advances in the number and quality of KS-like feature extraction procedures, the number of sufficiently sophisticated and diverse KSs that are needed to implement a general purpose computer vision system is not readily available. Due to this lack of resources, pursing an action in EHCVIS is accomplished by simulating the tasking or invocation of KSs. Let us describe this process.

5.3.2 Simulating the invocation of KSs. EHCVIS has a pool of nineteen KSs that are capable of providing a variety of information. A subset of these KSs are typically called low-level feature extraction processes. In aggregate, these KSs can express opinions about a region's texture, spectral properties, two-dimensional spatial relationships to other regions, and its polygonal shape. The remaining subset of KSs are typically considered higher-level sources of information, called object KSs. The object KSs, in aggregate, can express opinions about the presence or absence, in a region, of visual entities such as roofs, houses, grass, tree crowns, and so on. Objects can be viewed, in a sense, as features of more complex scenes such as residential neighborhood scenes, farm scenes, and city scenes. Just as objects exhibit certain shape, spectral, and texture features, so can complex scenes exhibit features such as, houses, roofs, grass, and roads.

Every region in a segmented image that the system is expected to interpret contains nineteen mass functions, one for each KS. Each mass function represents a subjective estimate of the best opinion the corresponding KS can possible convey if it were asked to extract feature information from some region under examination. These subjective mass functions are generated and stored in an image data-base prior to interpretation.

These mass functions were derived by evaluating an empirical and or theoretical analysis of the algorithms it is expected a real KS will use when forming opinions. This evaluation process was repeated for each KS and for all the regions that were known to contain a specific object the system might be expected to discern. For regions containing multiple objects a different set of statistics would be computed and as a consequence a different mass function would have been generated and stored in the image data-base. The details of the process are explained in [44].

Forming the best opinion a KS might convey is not the objective of our simulation. Rather, the data-base of subjective mass functions is required in order to begin the simulation process that can be summarized in the five steps shown in Figure 5.

We recall to the readers attention that one of our our motives for using the DS theory is due to its increased ability to deal with limited evidential information. To the extent a KS's opinion is modeled with respect to these three characteristics of information, we will be able to evaluate the viability of the DS theory in our task domain. Therefore, what we are truly simulating is the degradation of a KSs opinion (i.e., mass function) with respect to certainty, precision, and accuracy.

The degradation process can be modeled as a function, D, of five parameters, Θ_Q, a mass function M, a certainty factor (cer), a precision factor (pre), and an accuracy factor (acc). In equation form:

$$D\,(\Theta_Q,\ M,\ cer,\ pre,\ acc) = M', \tag{27}$$

where M' is the degraded mass function. The cer, pre, and acc parameters specify the extent to which M is to be degraded with respect to the three characteristics of information. Let us provide our intuition and computational definition of how a mass function M is degraded to M'.

A KS should attribute a greater portion of its unit mass to a proposition, say $p \subset \Theta_Q$, the more certain it is about the truthfulness of that proposition. Conversely, a proportionately smaller amount of mass should be attributed to p if a KS is less certain about the proposition's truthfulness. In our simulation of the degradation in a KS's mass function, the *cer* parameter is used to determine the degree to which a KS's opinion should be made less certain – i.e., how much to reduce the amount of mass that has been attributed to a proposition. This is reflected in the following equation. For $0 \leq cer \leq 1$, the degree to which a KS's mass function M is to be degraded with respect to the certainty of a proposition p:

$$M'(p) = \begin{cases} cer * M(p), & \text{for } p \subset \Theta_Q; \\ M(\Theta_Q) + \sum(1 - cer) * M(p), & \text{for } p = \Theta_Q. \end{cases} \tag{28}$$

Notice that the amount of mass that was originally attributed to total ignorance (i.e., Θ_Q) is increased by the sum of the mass that was "taken" away from proper subsets of Θ_Q. Doing so insures that the constraint in equation 11 remains satisfied.

A KS that attributes a non zero amount of mass to a singleton in a frame is said to be expressing the most precise opinion possible with respect to that frame. Conversely, a KS that attributes a non zero amount of mass to Θ is expressing the least precise opinion possible. That is, attributing any non zero mass to a set of cardinality one is expressing a very precise opinion. And the precision of that opinion decreases as the cardinality of that set increases. The *pre* parameter in our simulation process controls the cardinality of a proposition p – i.e., its corresponding subset of Θ_Q. For $0 \leq pre \leq 1$, the degree to which a KS's mass function M is degraded with respect to a proposition p is given by:

$$M'(p') = p \bigcup ran\text{-}set\text{-}gen(\Theta_Q - p,\ pre), \tag{29}$$

where for some set, $s \subset \Theta_Q$, *ran-set-gen*$(s,\ pre)$ returns a random set $s' \subset \Theta_Q$ of cardinality $(1 - pre) * |s|$. Since our system does not have any particular knowledge about

how a KS's opinion becomes less precise the *ran-set-gen* function randomly selects the propositions to include in s.

A KS is said to be expressing an inaccurate opinion if it attributes a non zero amount of mass to any proposition that is not true with respect to the available evidence. Furthermore, the more mass it attributes to a false proposition the more it is in error. In our simulation scheme, for $0 \leq acc \leq 1$, the degree to which a KS's opinion is accurate, the degradation of that opinion with respect to accuracy is given by:

$$\begin{aligned} M'(\neg p) &= (1 - acc) * M(p); \\ M'(p) &= acc * M(p). \end{aligned} \tag{30}$$

We argue that the *cer*, *pre*, and *acc* parameters allow our simulator to model most, if not all, of the ways opinions might vary when expressed in terms of propositions in a frame of discernment. By specifying these three parameters, it is possible to characterize any degradation in an opinion that might be expressed by any real or imaginary KS.

Returning to Figure 5, the process of simulating the invocation of KSs involves first retrieving, from an image data-base, a mass function for each KS that is invoked. Next, each of these mass functions is degraded with respect to the *cer*, *pre*, and *acc* parameters. The result is a set of degraded mass functions that are then pooled using Dempster's rule. Finally, the consensus opinion formed by Dempster's rule is input to an inference engine that updates the *Spt* and *Pls* of propositions in Θ_Q.

5.3.3 Long Term Memory. The frame Θ_Q is the system's relatively static representation of the world and domain knowledge that is needed to "understand" images – commonly called long term memory (LTM) [41], [23]. LTM is the representation of the semantic relationship between observable features and the visual entities the system

might try to discern. In EHCVIS, the set of label hypotheses (i.e., propositions) in LTM is defined to be:

$$\begin{aligned}\Theta_Q = \{&\texttt{tree-crown-scene, sky-scene, shutters-scene, roof-scene,}\\ &\texttt{road-scene, residential-scene, house-scene, bush-scene,}\\ &\texttt{Puffton-house-scene, Griffith-house-scene, Brown-house-scene,}\\ &\texttt{front-wall-scene, side-wall-scene, grass-scene}\}.\end{aligned} \tag{31}$$

The propositions `Puffton-house-scene`, `Griffith-house-scene`, and `Brown-house-scene` represent particular house scenes that are associated with a particular individual or place. This is in contrast to a generic house scene as represented by the proposition `house-scene`. The reason `-scene` appears as a suffix to the above propositions such as `roof-scene`, `sky-scene` is that for a particular image or subset of regions in an image, the system might only be observing these objects. How a conjunction of label hypotheses that are not explicitly represented in Θ_Q can be instantiated in STM is explained in [44].

There are five feature spaces, F_1 through F_5, any subset of which might be used to partition Θ_Q. Enumerating all of these feature spaces and their corresponding sets, $\mathcal{F}_1$ through $\mathcal{F}_5$, of feature propositions would be excessive for this paper. However, we shall enumerate a subset of LTM to help make the remaining discussion more lucid.

The following three feature spaces are associated with the indicated types of visual information that might be used to discern propositions in Θ_Q:

F_1 **Objects:** such as tree crowns, roofs, roads, and so on;

F_2 **Spectral:** such as grass green, sky blue, road black, road grey, and so on;

F_3 **Texture:** such as highly textured grass, smoothly textured sky, and so on.

The following is an enumeration of some of the feature propositions in the above feature spaces:

$\mathcal{F}_1$ has-sky-as-part, has-grass-as-part, has-walls-as-part, ...

$\mathcal{F}_2$ has-sky-blue-as-part, has-grass-green-as-part, ...

$\mathcal{F}_3$ has-bush-angular-line-density-as-part,
has-house-angular-line-density-as-part,
has-sky-angular-line-density-as-part, ...

The reason the spectral feature proposition has-sky-blue-as-part, for instance, specifies the object sky is due to two reasons. The first is that there is no universally standard quantification of the color blue in an image that was produced by some uncalibrated photographic process. Such photographic images are commonly used to generate a digitized image of the original scene. The second is that without this calibrated information, the only way to currently capture some measure of "blueness" is to sample a collection of regions in images that contain only blue skys. As a consequence, what one has actually measured is not blueness, rather sky-blueness. And in a similar fashion, one can only measure grass-green, grass-blueish-green, road-grey, and so on.

For each $\mathcal{F}_i$ we must construct a χ_i to partition Θ_Q. Again, a complete enumeration is excessive, however we shall list a subset of the χs actually defined in EHCVIS. For $\mathcal{F}_1$:

$$\begin{aligned}
\chi_1(\texttt{has-sky-as-part}) =& \{\texttt{sky-scene, road-scene,} \\
& \texttt{residential-scene, Puffton-house-scene,} \\
& \texttt{Griffith-house-scene, Brown-house-scene}\}; \\
\chi_1(\texttt{has-grass-as-part}) =& \{\texttt{grass-scene, road-scene,} \\
& \texttt{residential-scene, Puffton-house-scene,} \\
& \texttt{Griffith-house-scene, Brown-house-scene}\},
\end{aligned} \tag{32}$$

and so on. And finally for $\mathcal{F}_3$:

$$\begin{aligned}
&\chi_3(\texttt{has-tree-crown-angular-line-density-as-part}) = \\
&\qquad \{\texttt{tree-crown-scene, road-scene,} \\
&\qquad \texttt{residential-scene, Puffton-house-scene,} \\
&\qquad \texttt{Griffith-house-scene, Brown-house-scene}\}; \\
&\chi_3(\texttt{has-grass-angular-line-density-as-part}) = \\
&\qquad \{\texttt{grass-scene, road-scene,} \\
&\qquad \texttt{residential-scene, Puffton-house-scene,} \\
&\qquad \texttt{Griffith-house-scene, Brown-house-scene}\},
\end{aligned} \tag{33}$$

and so on.

The output of the simulation process is a set of degraded mass functions. These mass functions are then combined by Dempster's rule to form a consensus about which label hypothesis is appropriate to associate with the current region of interest. The result of applying this rule is input to an inference engine that updates the Spt and Pls of propositions in LTM. After the updating is completed the results are evaluated in phase four.

5.4 PHASE FOUR:

5.4.1 **Evaluating LTM.** The evaluation of LTM and the state of the system up to this point can be characterized in four steps.

STEP 1: The first step involves determining if there was sufficient conflict between KSs to justify verifying the KSs. The details of the verification process in EHCVIS is complex – see [44]. However, if the KSs have been verified or the conflict they generate is below some threshold, then proceed to **Step 2.**

STEP 2: In the second step, the system tries to determine if the consensus opinion that was formed over LTM is sufficient to justify instantiating a label hypothesis – i.e., the Spt of one or more label hypotheses is above some minimum threshold. If so, then instantiate the hypotheses in a representation called short term memory (STM). Then place on the goal stack the goal of looking for objects that can possibly coexist with the object hypotheses that were previously instantiated in STM, and then go to **PHASE ONE**, else go to **Step 3.**

STEP 3: If the system reaches this step, then it is trying to reduce the amount of ambiguity, dissonance, or ignorance for a subset of label hypotheses in Θ_Q. If the maximum number of attempts to instantiate a hypothesis has not been reached, then the goal of obtaining additional information about the currently best label hypotheses is put on the goal stack. Then the system proceeds to **PHASE ONE** else to **Step 4.**

STEP 4: This step is reached if EHCVIS has exhausted the maximum number of attempts to instantiate a hypothesis. At this point, the goal of identifying objects the system has the best chance of discerning is put on the goal stack. Then the system proceeds to **PHASE ONE** else the interpretation process is terminated if the maximum number of attempts at interpreting the entire image is exceeded.

It is clear that the evaluation phase of our system plays an important role in controlling the interpretation task. Indeed, the selection of goals and their constraints is dependent on factors that our system does not yet taken into account. Some of the limitations and consequences of this are discussed in [44].

This completes the discussion of EHCVIS and how both the DS and ER technologies have been integrated into the system's mechanisms for reasoning about the control of the interpretation process and reasoning about the visual entities it is expected to perceive.

Next we shall briefly discuss the interpretation experiments that were conducted and summarize their results.

6 INTERPRETATION EXPERIMENTS

There are several objectives of the research and experiments that are reported here and in [44]:

1) to demonstrate that certain types of incompletenesses in LTM can be detected when certain "evidential measures" and verification procedures are employed;

2) to demonstrate that the system tends to degrade smoothly as the quality of the information it must reason from becomes less certain, precise, and accurate, and;

3) to demonstrate that a system's performance is improved (i.e., fewer resources are used, better interpretations, or a combination of the above) as more evidential-based control strategies are used.

In this paper, we shall begin to emphasize the later two objectives by describing our experimental design, method, and results.

Over one hundred and forty interpretation experiments were conducted on three digitized and segmented color monocular images of outdoor natural scenes that are similar to Figure 1. Each experiment involved selecting a value for each of the three degradation parameters and then tasking EHCVIS to interpret the image. For instance, we typically started a series of experiments with $cer = pre = acc = 1$, then after the system did its best at completing the interpretation task, the degradation parameters we set to $cer = .9$, and $pre = acc = 1$, then $cer = .8$, and $pre = acc = 1$, ..., $cer = 1, pre = .9, acc = 1$, and so on until $cer = pre = acc$ equaled $.4$ or $.5$. This sequence of degradation in the three parameters was conducted twice for each image. That is, once without using any control strategies to establish a baseline level of performance. The remaining times the system was allowed to employ various combinations of control strategies. This allowed

us to compare performance between the use of various control strategies and no control strategy.

For each experiment, the KSs in the system would be most certain, precise, and accurate when *cer*, *pre*, and *acc* equaled one, respectively. Conversely, the KSs became less certain, precise, and accurate as *cer*, *pre*, and *acc* approached zero, respectively. Experiments were not conducted with parameter values for which it was clear the system would not be capable of interpreting the image.

Several metrics were used to measure the system's performance, one being the number of correctly instantiated regions. But before we discuss the system's performance, let us briefly annotate a portion of the system's attempt at interpreting the image and its segmentation in Figures 1 and 2 respectively.

6.1 ANNOTATED INTERPRETATION TASK

The portion of an interpretation experiment described here is intended to demonstrate one important point. That by taking advantage of the additional information the DS makes readily available, our system was able to identify objects that were previously undiscernible when this information was unavailable.

Consider Figures 6 through 12. At a point early in the process of trying to interpret region R_{14} of the segmented image in Figure 6, the system was unable to disambiguate whether that region is the `side-wall-scene` or the `front-wall-scene` of the house in the image. During this experiment, the *cer*, *pre*, and *acc* parameters were set to 1, .7, and 1, respectively. That is, all the KSs were as certain, and accurate as possible, however, they were made 30% less precise than they could be. Figure 7 shows that after reaching the maximum number of attempts to interpret the region the system remained most ambiguous about the `side-wall-scene` and `front-wall-scene` label hypotheses. Thus, without using a control strategy to help resolve this ambiguity the object in region R_{14} was not identified. As illustrated in Figure 8, the only way the system would be

capable of resolving this ambiguity would be to obtain information that distinguishes `side-wall-scenes` from `front-wall-scene`.

As mentioned earlier, as part of its specification, each KS can typically only attribute mass to a subset of the feature propositions in some feature space. In Figure 9, we see that KS_6 can attribute mass for or against the `has-walls-as-part`, `has-side-walls-as-part`, and `has-front-wall-as-part` feature propositions. In contrast, we also see in Figure 9 that the only feature proposition KS_1 can attribute mass for or against is `has-house-as-part`. Therefore, at this point in the interpretation, KS_6 appears to be better suited for resolving the ambiguity of current interest.

However, by allowing the system to use its ambiguity resolving control strategy, we can begin to see in Figure 10 how the system might be able to label R_{14}. In Figure 10, the two control strategies used in this experiment are underlined at the top of the figure. The two CKSs that are responsible for measuring *Igr*, and *Amb* are $\texttt{CKS}_2$ and $\texttt{CKS}_4$ respectively. The set of possible actions the system might take (i.e., Θ_A) is enumerated in the list under the title "`Pruned *action-prop-names.`" These alternatives were the same as those available to the system when the above control strategies were not used. After both $\texttt{CKS}_2$ and $\texttt{CKS}_4$ have made their respective measurements, they construct mass functions that reflect their opinion about which alternatives they believe is the best to pursue. We see that $\texttt{CKS}_2$ believes very strongly that taking those actions that invoke KS_6 is more appropriate than taking those actions that do not invoke KS_6. Likewise, $\texttt{CKS}_4$ believes, almost as strongly, the same as $\texttt{CKS}_2$. We see in Figure 11 the result of pooling these two opinions over Θ_A. That is, the consensus opinion strongly indicates that taking either action a_1 or a_2 is appropriate because they result in tasking KS_6, which has the best chance of resolving the ambiguity of concern. The results of the system actually pursuing a_2, which was randomly chosen from $\{a_1, a_2\}$, is illustrated in Figure 12. The opinion of KS_6 was such that it supported a proposition that distinguished `front-wall-scene` from `side-walls-scene` to a degree that allowed the system to instantiate the `front-wall-scene` label hypothesis for R_{14}.

There was no guarantee that KS_6 would have helped discern the propositions of interest. Rather, among the available KSs it was the most likely to provide the system with the needed information. The ambiguity control strategy biased the system to take those actions that were most likely to result in obtaining the desired information.

The portion of an actual experiment just presented illustrates how EHCVIS uses the DS and ER technology to accomplish two major interpretation tasks that were described earlier in this paper: 1) to reason about what label hypotheses to assign to regions in an image and; 2) to decide how its limited resources should be utilized in order to complete the image interpretation task. A number of experiments were conducted using a variety of control strategies (e.g., reliability of KSs, dissonance resolving control strategies, and so on) in conjunction with various combinations of degradation parameter values.

The results of the experiments we have conducted can be and are presented in a number of ways, see [44]. Here, we shall summarize these results with respect to one performance measure: the number of correctly instantiated regions. In addition, the results presented in this section are with respect to experiments on the image in Figure 1. However, the results for the remaining two images are similar to those presented here.

In summary, when all the KSs were as certain, precise, and accurate as possible, (i.e., $cer = pre = acc = 1$), and no control strategies were used, the system was able to correctly label approximately 90% of the regions it examined. When all the KSs were as certain, precise, and accurate as possible and the system was allowed to use any number of control strategies, the system was able to correctly label approximately 91% to 92% of the regions examined. This suggests that "evidential control strategies" do not significantly improve a system's performance when its sources operate at optimum levels. However, as the KSs became less certain, but remained as precise, and accurate as possible, and no control strategies were used, the system was able to correctly label only approximately 23% of the regions examined for a 40% decrease in certainty. But when the system was allowed to use any number of control strategies, it was able to correctly label as many as 70%

of the regions examined for the same 40% decrease in just the certainty of a KS's mass function. The level of performance was qualitatively the same when the mass functions of KSs were degraded with respect to just accuracy or just precision. The degree to which the system's performance was improved when the mass functions of KSs were degraded with respect to certainty, precision, *and* accuracy was not as dramatic as the results just described indicate. However, the improvement that was noticed was significant enough to justify using these "evidential" control strategies.

In short, the results indicate that although taking advantage of the information the DS theory provides does not significantly improve a KBS's performance when its perceptions are near perfect. The benefits of using such information becomes obvious as the quality of a KBS's perceptions degrade. That is, the degradation of the system's performance is significantly delayed.

7 REASONING IN COMPUTER VISION SYSTEMS: RELATED WORK

There are some important similarities and differences between our approach to reasoning from limited evidential information and that used by others – see for instance Nagao and Matsuyama [28], Brooks [4], Peter Selfridge [33], Kenneth Sloan [37], Thomas Garvey [13], Hanson and Riseman [19], and Levine and Shaheen [24].

The object recognition portion of Nagao's and Matsuyama's system uses, in part, a boolean approach to reasoning about the perceptions of its KS-like feature extraction processes. Their approach is similar to ours in that semantic knowledge about objects are represented in terms of object-features that can and cannot coexist – e.g., see table 6.1 in [28]. The approaches differ in how beliefs about the presence or absence of object-features in a region of interest are represented, pooled, and how inferences are drawn from these beliefs. In Nagao's system beliefs about the presence or absence of any particular object feature in some region of interest is represented in a Boolean "yes" or "no" manner. This boolean decision is made in the source (e.g., KS) that must express its beliefs. These beliefs

are then pooled in a logical fashion to infer which label hypothesis should be instantiated for the region under examination. However, the difficulties of reasoning in a Boolean fashion have been discussed in [25], [26]. In contrast, KSs in our system express, on a continuous scale, their partial beliefs about the presence or absence of object-features in a region. And we have previously pointed out the benefits of providing KSs with this flexibility.

The work of Brooks [4], Peter Selfridge [33], Kenneth Sloan [37], Thomas Garvey [13], Hanson and Riseman [19], Levine and Shaheen [24], Yakimovsky and Feldman [46], and Zucker [48] for the most part employ mechanisms that are probabilistic, Boolean, or an ad hoc variant thereof for pooling beliefs and drawing inferences. Therefore, it is difficult if not impossible for their systems to take advantage, in a nice formal way, of evidential measures such as the amount of ignorance, dissonance, ambiguity, decisiveness and so on a proposition might exhibit. This work suggests that the performance of their systems might improved if they take advantage of such evidential information.

8 SUMMARY

In this paper, we have discussed research on the application of both the Dempster-Shafer theory and the concept of evidential reasoning in order to begin addressing several problems that KBSs must deal with. Our domain of application was knowledge-based computer vision. The DS theory and concept of ER is the foundation of a developing framework for knowledge-based systems, such as general purpose computer vision systems, that must reason in complex domains about both their perceptions and the actions they might pursue in order to understand their environment. Some results from a large number of interpretation experiments were summarized to highlight a few of the benefits of employing these technologies in a large scale knowledge-based system. That is, by using previously unavailable information such as the amount of dissonance, ignorance, and or ambiguity a label hypotheses exhibits, the system was able to correctly label a significantly greater number of regions in an image.

Despite the progress of this research, there remains a significant number of problems to address with respect to the technology we have explored and its use in knowledge-based systems. For instance, although the DS theory has relieved us from the burden of specifying complete probability models, a formal theory for generating mass functions remains unavailable. We believe that this later problem is more tractable than the former. Another concern is, given the independence requirement of Dempster's rule, is there a formal model by which dependencies can be automatically accounted for in a frame of discernment? And finally, but not the least of which is, the lack of a computational theory for the integration of "fuzzy-based" approaches to uncertain reasoning with the theory of belief functions [47].

9 ACKNOWLEDGMENTS

We would like to thank Joey Griffith, currently a graduate student in the Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts 01003, for his unpublished work on the algorithms that produced the segmentations from the images used in this work. In addition, we would like to thank Allen Hanson, John Lowrance, Charles Randall, and Edward Riseman for their assistance in completing this research, and Stephen Lesh, Thomas Strat, and Enrique Ruspini for their constructive comments.

The work reported here was supported in part by the Defense Advanced Research Projects Agency (DARPA) under Contract No. N00014-81-C-0115, (DARPA) under Contract No. N00014-82-K-0464, by the Navy Electronic Systems Command under Contract No. N00039-83-K-0656 (SRI Project No. 6486) and Contract No. DAAB07-84-C-F092 (SRI Project No. 7845), and while enrolled in the Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, by Air Force Grant No. AFOSR-85-0005, and by Air Force Contract No. F49620-83-C-0099. The views and conclusions contained in this document are those of the author and should not

be interpreted as representing the official policies, either expressed or implied, of SRI International, DARPA, the Navy, the U.S. government, or the University of Massachusetts.

Figure 1.

A MONO-CHROMATIC RENDERING OF A TYPICAL STATIC 2-D COLOR IMAGE OF AN OUTDOOR NATURAL SCENE.

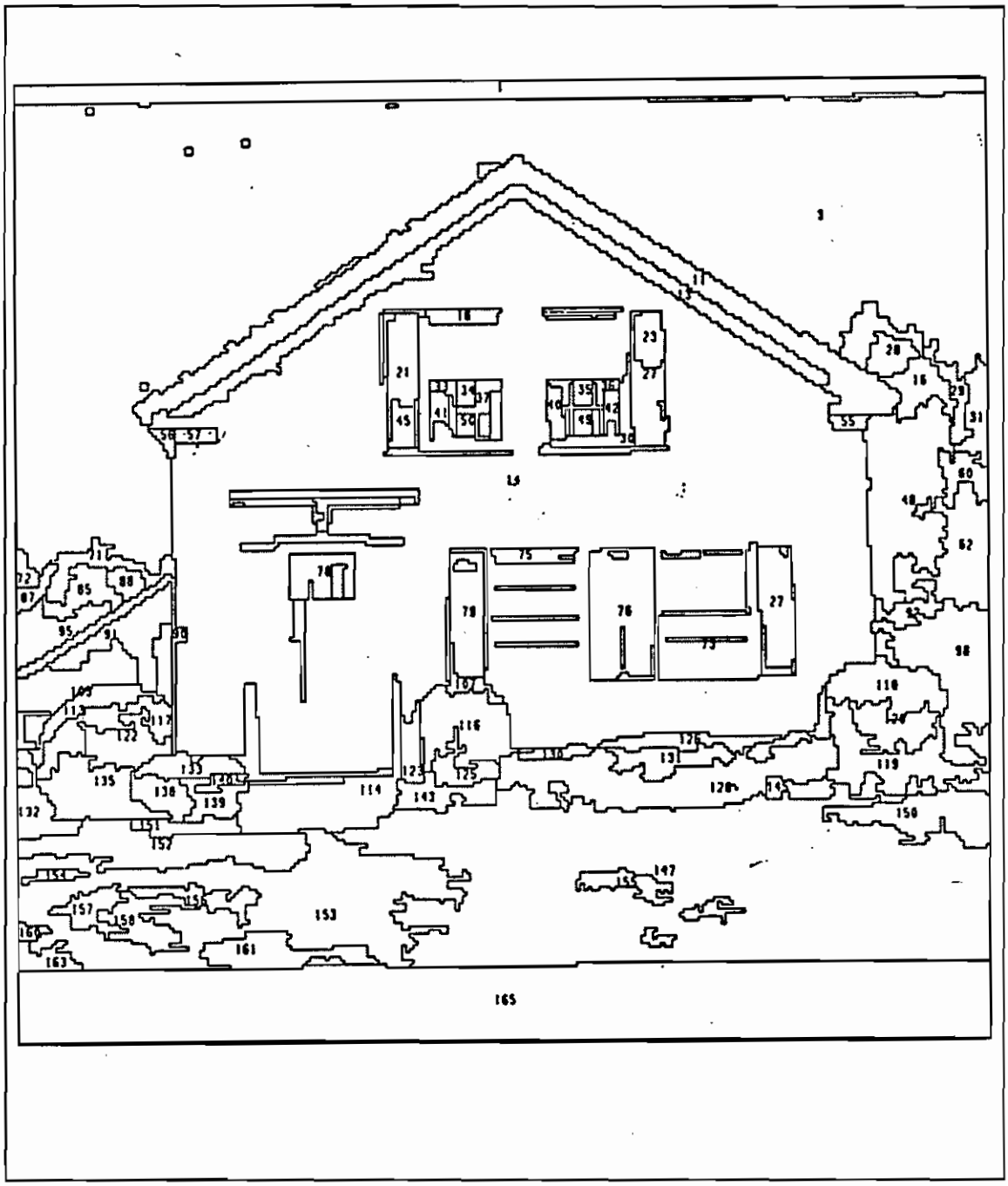

Figure 2.

AN EXAMPLE SEGMENTATION OF THE IMAGE IN FIGURE 1.

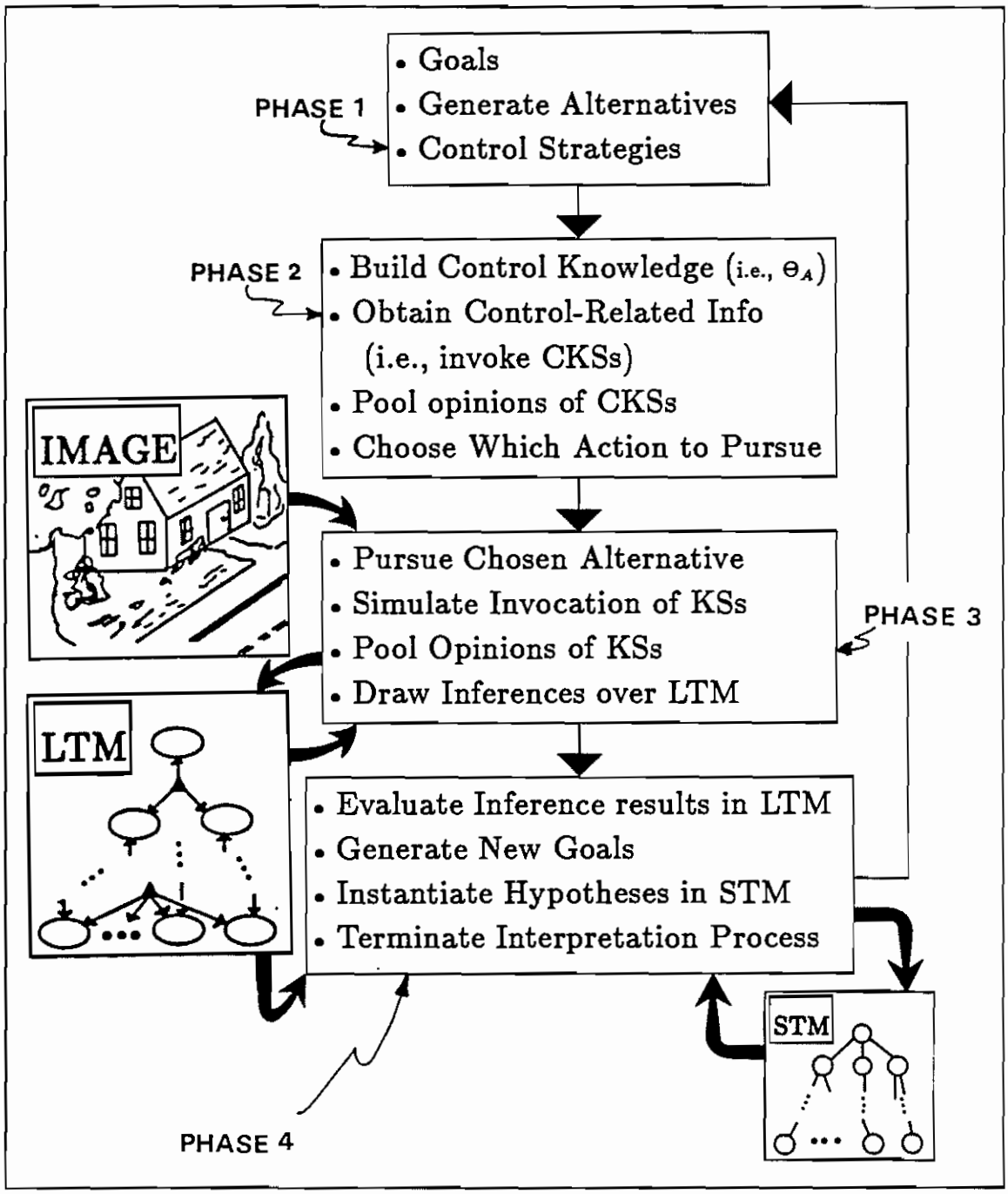

Figure 3.

A SYSTEM FLOW DIAGRAM OF EHCVIS

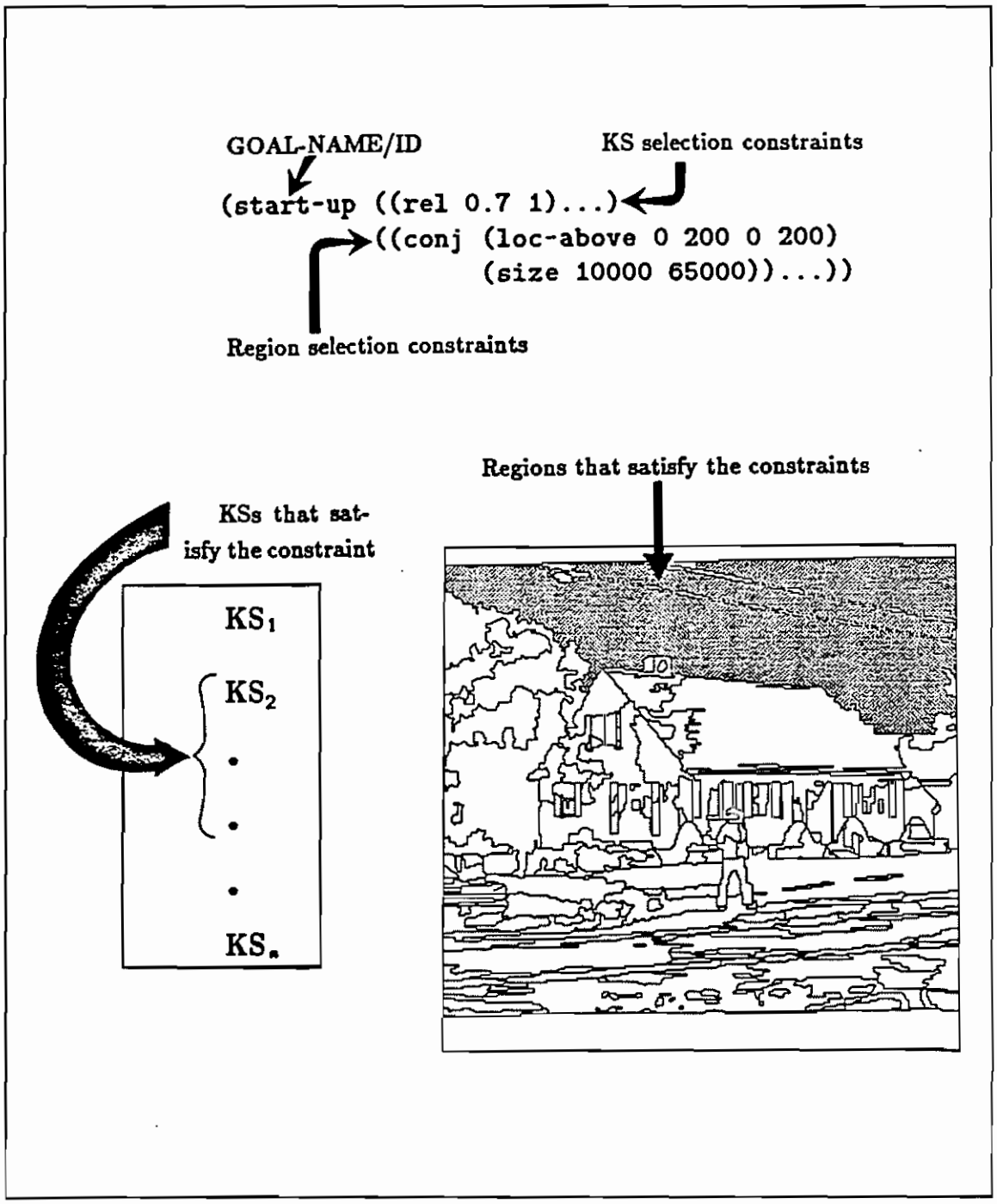

Figure 4.

EXAMPLE GOAL AND CONSTRAINTS.

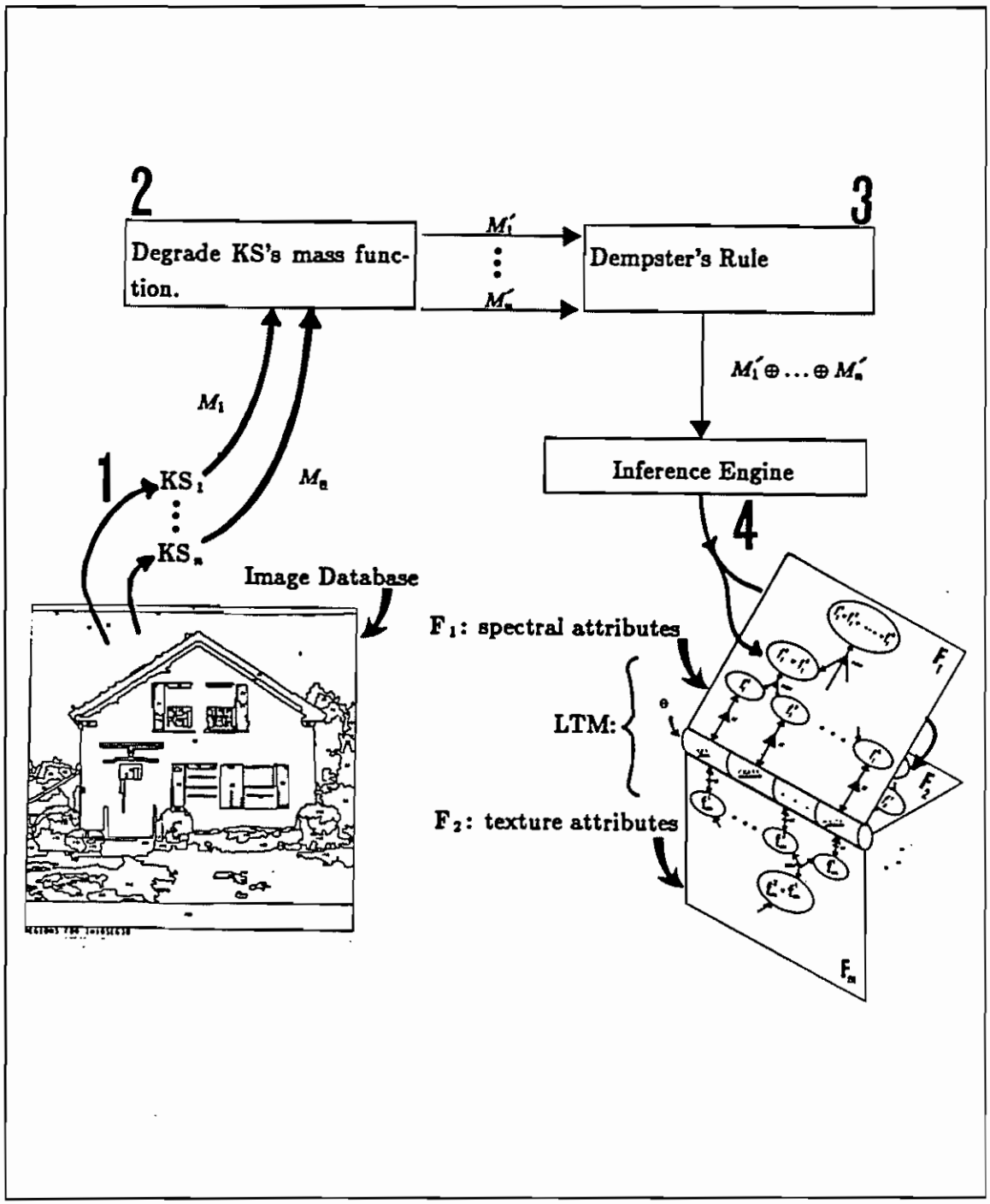

Figure 5.

SIMULATING KS INVOCATIONS.

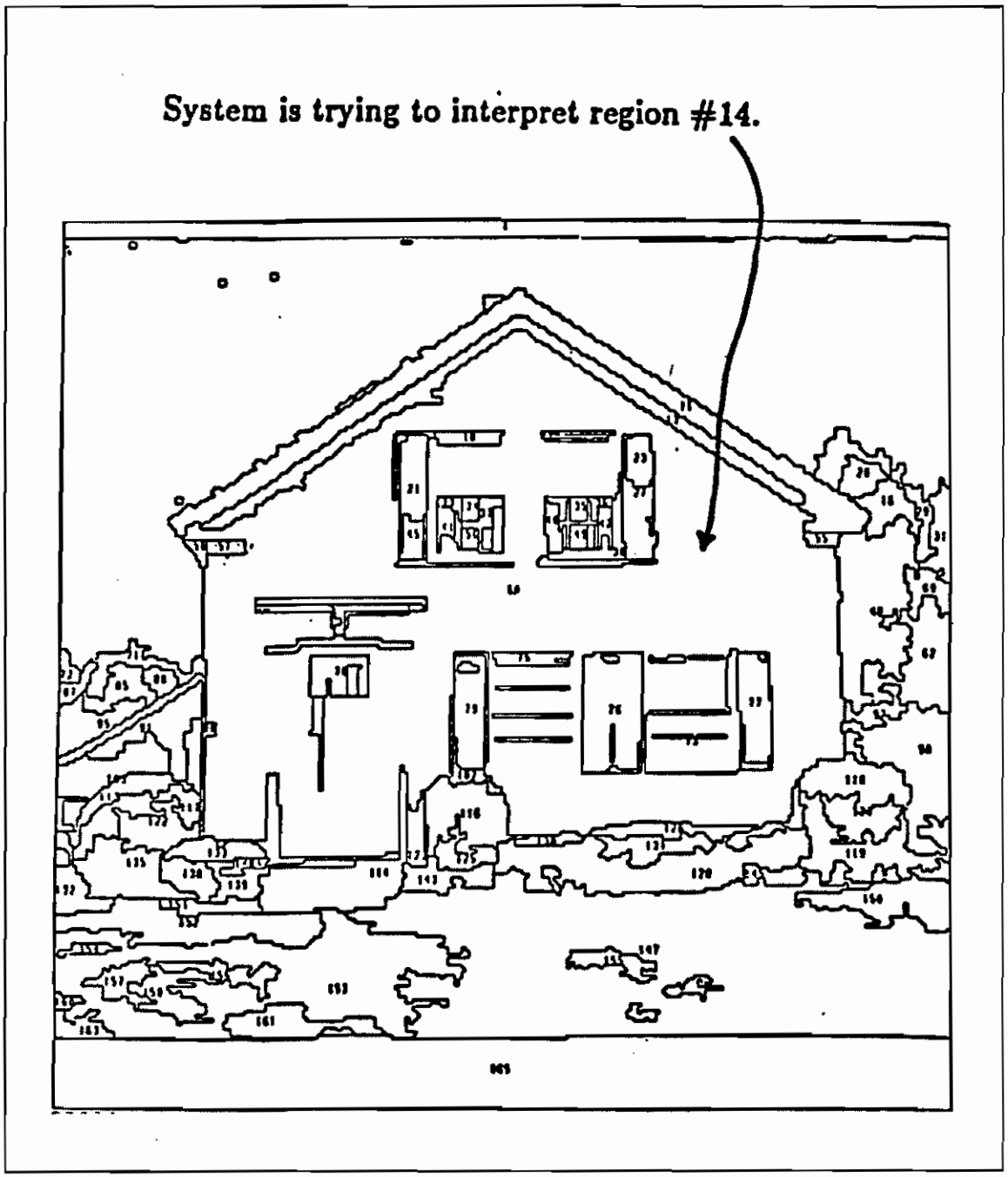

Figure 6.

SEGMENTATION OF AN IMAGE THE SYSTEM IS TRYING TO INTERPRET.

Without ambiguity control strategy:

```
precisenesŝ parameter value = .7

invoke~KS80&KS70&KS30~R14  ==> ((KS80 KS70 KS30) (R14))

==========================================================
                LTM inference results
==========================================================
tree-crown-scene      [0.0 , 0.0]    !*------------------!
sky-scene             [0.0 , 0.0]    !*------------------!
side-walls-scene      [0.0 , 1.0]    !********************!
shutters-scene        [0.0 , 0.0]    !*------------------!
roof-scene            [0.0 , 0.0]    !*------------------!
road-scene            [0.0 , 0.0]    !*------------------!
puffton-house-scene   [0.0 , 1.e-3]  !*------------------!
house-scene           [0.0 , 1.e-3]  !*------------------!
griffith-house-scene  [0.0 , 1.e-3]  !*------------------!
grass-scene           [0.0 , 2.5e-2] !*------------------!
front-wall-scene      [0.0 , 1.0]    !********************!
bush-scene            [0.0 , 0.0]    !*------------------!
brown-house-scene     [0.0 , 0.0]    !*------------------!
a-road-scene          [0.0 , 0.0]    !*------------------!

==========================================================
```

Figure 7.

CURRENT STATE OF INTERPRETATION PROCESS WITHOUT USING ANY CONTROL STRATEGY.

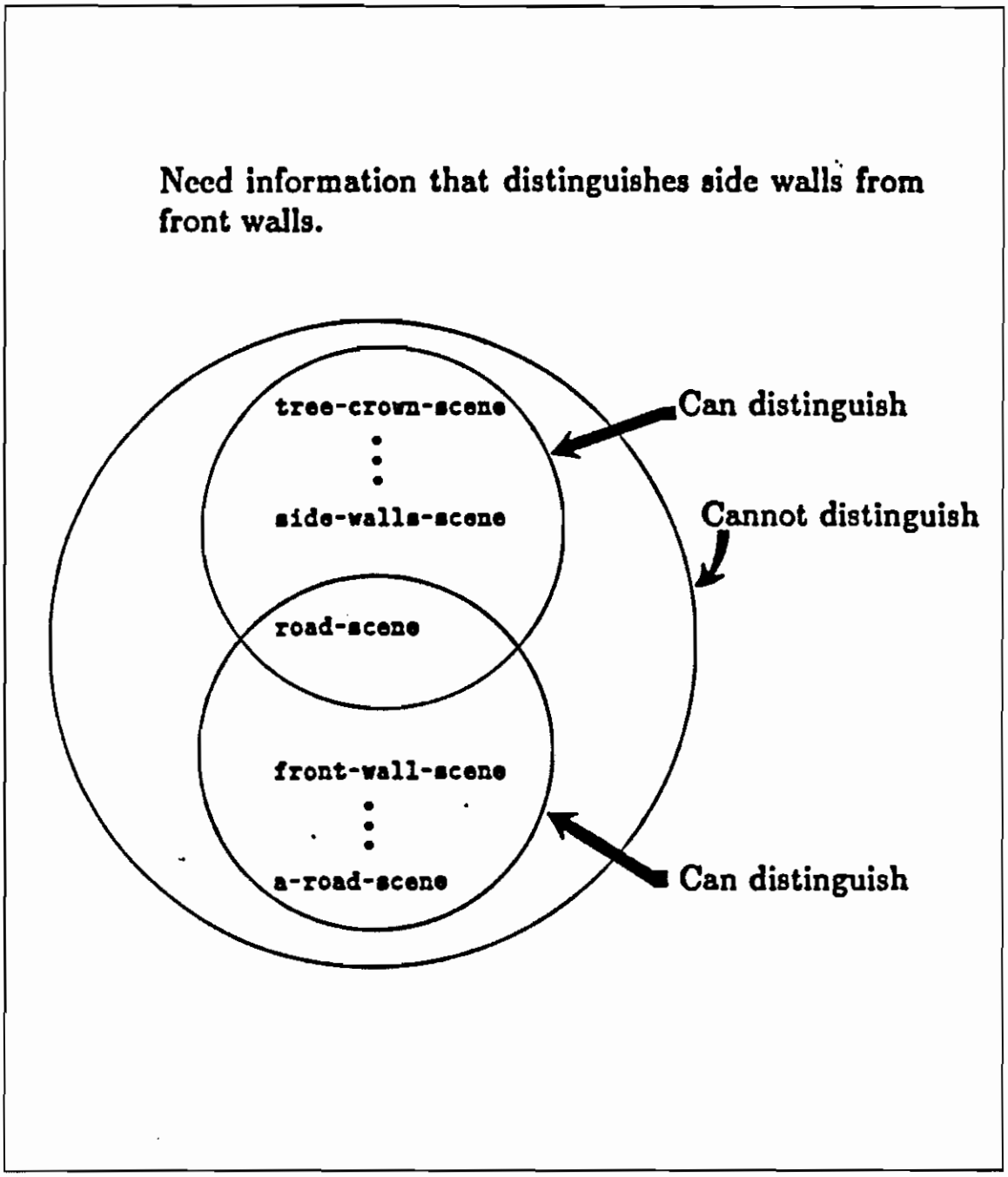

Figure 8.

ILLUSTRATION OF INFORMATION NEEDED TO DISCERN AMBIGUOUS LABEL HYPOTHESES.

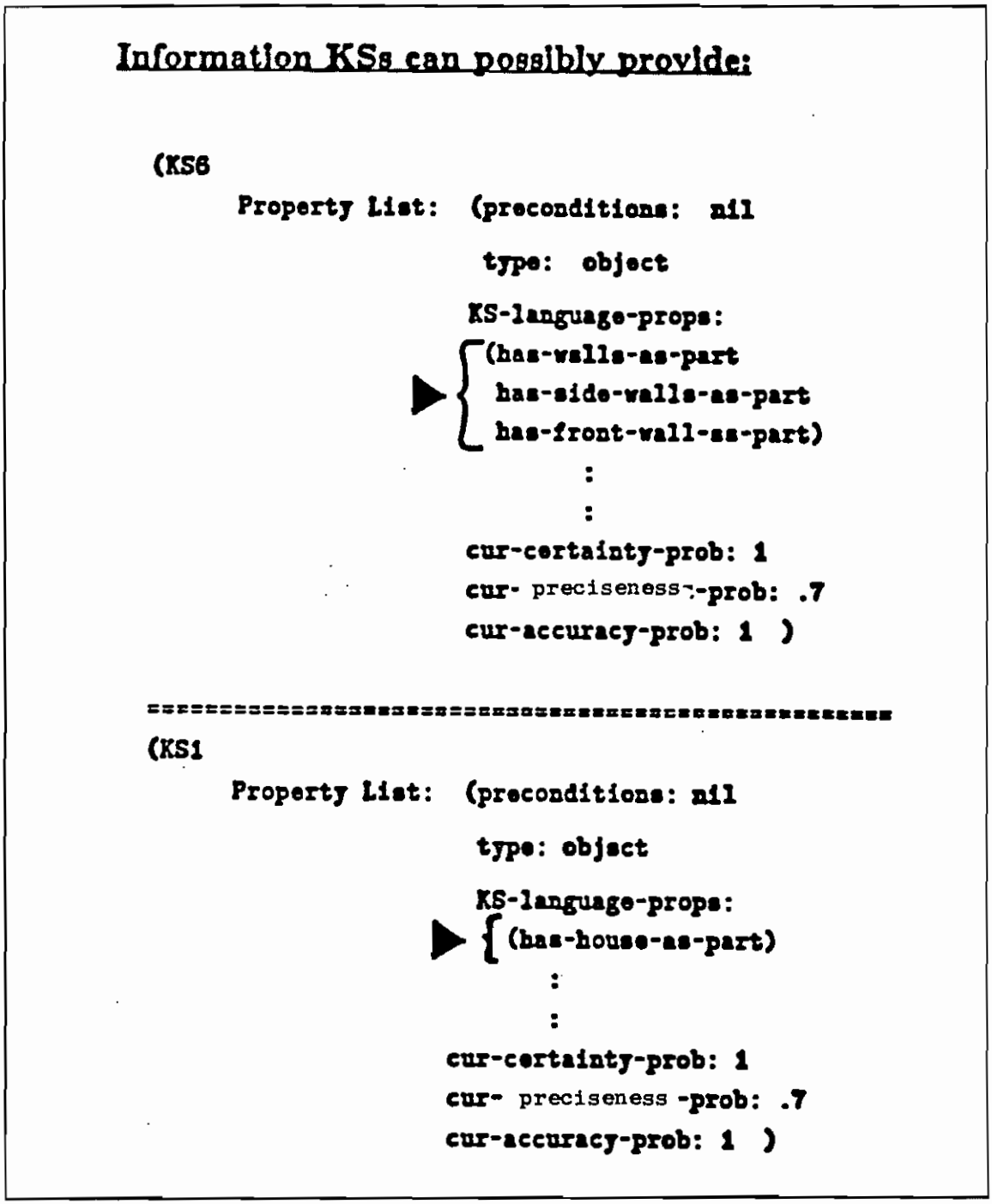

Figure 9.

KS SPECIFICATIONS.

Control strategies & CKS mass-functions:

```
TYPE OF INFORMATION REPORTED: *strategies

(most-igr-about-prop best-ambiguity-resolving-ks-is)
=========================================

TYPE OF INFORMATION REPORTED: Pruned *action-prop-names

(invoke~KS70&KS30&KS80&KS90~R14 invoke~KS6&KS30&KS80&KS90~R14
 invoke~KS70&KS80&KS30~R14 invoke~KS6&KS70&KS30&KS90~R14)
=========================================

TYPE OF INFORMATION REPORTED: The mass functions returned by
                              the last invoked CKSs

((CKS2 <== most-igr-about-prop
  (((invoke~KS6&KS70&KS30&KS90~R14 invoke~KS70&KS80&KS30~R14
     invoke~KS6&KS30&KS80&KS90~R14
     invoke~KS70&KS30&KS80&KS90~R14)     .1)
   ((invoke~KS6&KS70&KS30&KS90~R14
     invoke~KS6&KS30&KS80&KS90~R14)      .9)))

 (CKS4 <== best-ambiguity-resolving-ks-is
  (((invoke~KS6&KS30&KS80&KS90~R14
     invoke~KS6&KS70&KS30&KS90~R14)      0.6667)

   ((invoke~KS70&KS30&KS80&KS90~R14 invoke~KS70&KS80&KS30~R14
     invoke~KS6&KS30&KS80&KS90~R14
     invoke~KS6&KS70&KS30&KS90~R14)      0.333))))
=========================================
```

Figure 10.

OPINIONS FROM IGNORANCE AND AMBIGUITY CKSs.

Combining CKS2's & CKS4's mass-functions:

Let:

a_1 = invoke-KS6&KS70&KS30&KS90-R14

a_2 = invoke-KS6&KS30&KS80&KS90-R14

a_3 = invoke-KS70&KS30&KS80&KS90-R14

a_4 = invoke-KS70&KS80&KS30-R14

$\Theta_A = a_1 \vee a_2 \vee a_3 \vee a_4$

F_1: best-ambiguity-resolving-ks

F_2: most-igr-about-prop

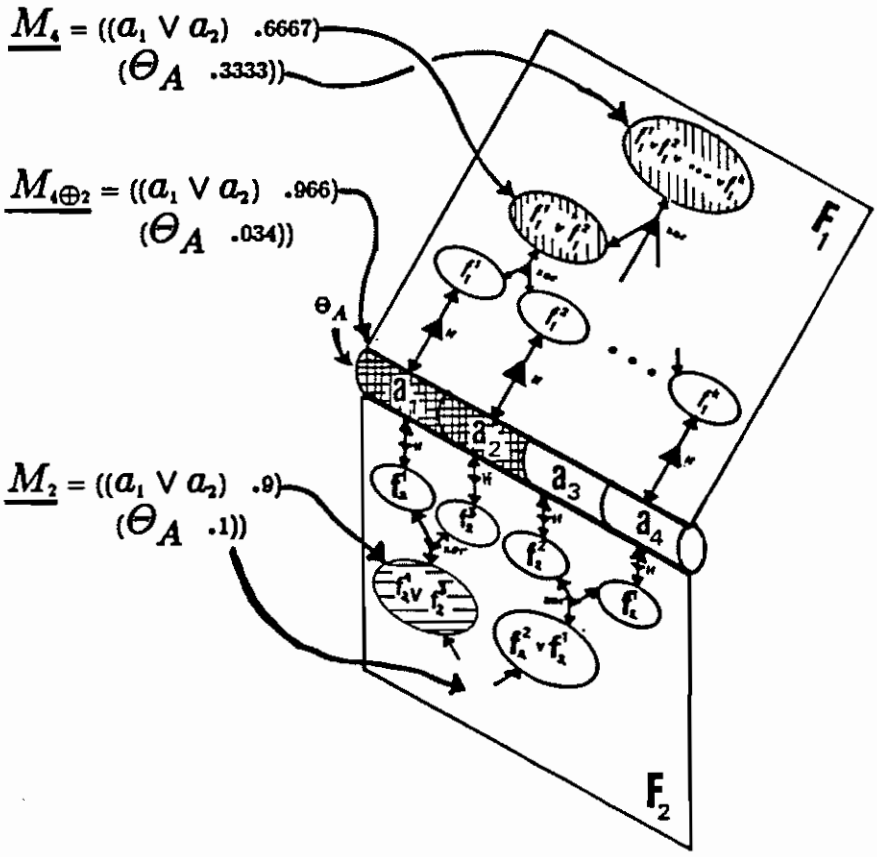

Figure 11.

POOLING THE OPINIONS OF THE IGNORANCE AND AMBIGUITY CKSs.

Results of taking action:

```
TYPE OF INFORMATION REPORTED: action system will take
invoke~KS6&KS30&KS80&KS90~R14  <== ((KS6 KS30 KS80 KS90) (R14))

==========================================================
                LTM inference results
==========================================================

tree-crown-scene      [0.0 , 1.e-3]   !*------------------!
sky-scene             [0.0 , 1.e-3]   !*------------------!
side-walls-scene      [0.0 , 0.3]     !******------------!
shutters-scene        [0.0 , 0.0]     !*------------------!
roof-scene            [0.0 , 1.e-3]   !*------------------!
road-scene            [0.0 , 1.e-3]   !*------------------!
puffton-house-scene   [0.0 , 3.e-3]   !*------------------!
house-scene           [1.e-3 , 4.e-3] !*------------------!
griffith-house-scene  [0.0 , 3.e-3]   !*------------------!
grass-scene           [0.0 , 1.e-3]   !*------------------!
front-wall-scene      [0.696 , 0.999] !-----------*******!
bush-scene            [0.0 , 1.e-3]   !*------------------!
brown-house-scene     [0.0 , 3.e-3]   !*------------------!
a-road-scene          [0.0 , 1.e-3]   !*------------------!

==========================================================

TYPE OF INFORMATION REPORTED:
instantiated hypothesis, instantiated regions,
((front-wall-scene) R14)
```

Figure 12.

RESULT OF USING AMBIGUITY CONTROL STRATEGY.

REFERENCES

[1] Barnett, Jeffrey A., in Seventh International Joint Conference on Artificial Intelligence, Proc. IJCAI, 868 (1981).

[2] Barnett, Jeffrey A., Combining Opinions About the Order of Rule Execution, working paper, Northrop Research and Technology Center, One Research Park, Palos Verdes Peninsula, California 90274, (1974).

[3] Binford, Thomas O., International Journal of Robotics Research, 1(1), 18 (1980).

[4] Brooks, R. A., Artificial Intelligence, 17(3), 285 (1981).

[5] Burns, Brian J., Hanson, Allen R., et. al., in Seventh International Conference on Pattern Recognition, Proc. 482 (1984).

[6] Davis, L. S., and A. Rosenfeld, Computer Vision Systems, Edited by Allen R. Hanson and Edward M. Riseman, Academic Press, New York (1978).

[7] Dempster, A. P., Annals of Mathematical Statistics, 38, 325 (1967).

[8] Dempster, A. P. and A. Kong, Belief Functions and Communications Networks, Final Report ARO Contract DAAG29-84-K-0128, Department of Statistics, Harvard University, Cambridge, Massachusetts, 02138, (1984).

[9] Duda, Richard O., et. al. in Proc. National Computer Conference, Proc. 40, 1075 (1976).

[10] Duda, Richard O. and Rene Reboh, et. al., A Computer-Based Consultant For Mineral Exploration, SRI Final Report for project # 6415, Artificial Intelligence Center, SRI International, Menlo Park, California 94015 September (1979).

[11] Fischler, M.A., Tenenbaum, J.M., et. al., Computer Graphics and Image Processing, 15, 201 (1981).

[12] Fischler, M.A. and H.C. Wolf, in Computer Vision and Pattern Recognition, Proc. CVPR, 351 (1983).

[13] Garvey, T.D., Perceptual Strategies for Purposive Vision, SRI International Tech. Note 117, Artificial Intelligence Center, Menlo Park, California, 94025, September (1976).

[14] Garvey, T. D., J.D. Lowrance, and M. A. Fischler, in Seventh International Joint Conference on Artificial Intelligence, Proc. IJCAI, 319 (1981).

[15] Ginsberg, M. L., in National Conference on Artificial Intelligence, Proc. AAAI, 126 (1984).

[16] Gordon, Jean, and Edward H. Shortliffe, Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Edited by Buchanan, B. G. and E. H. Shortliffe, Chap. 13, Addison-Wesley, Reading Massachusetts, (1985).

[17] Gordon, Jean, and Edward H. Shortliffe, Artificial Intelligence, A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space, to appear.

[18] Gorry, Anthony A System for Computer-Aided Diagnosis, Ph.D. dissertation, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Massachusetts, 02139, (1967).

[19] Hanson, Allen R., and Edward M. Riseman, Computer Vision Systems Academic Press, New York, (1978).

[20] Havens, W. and A. Mackworth, Computer 16(10), 90 (1983).

[21] Kong, A., Explaining a Paradox Using the Theory of Belief Functions, Tech. Note Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, (1983).

[22] Laws, K., Goal-Directed Textured-Image Segmentation, SRI Technical Note No. 334, Artificial Intelligence Center, SRI International, Menlo Park, California 94025, (1984).

[23] Levine, Martin D., Computer Vision Systems, Edited by Allen R. Hanson and Edward M Riseman, Academic Press, New York, (1978).

[24] Levine, Martin D. and Samir I. Shaheen, in Pattern Analysis and Machine Intelligence, Trans. PAMI-3(5), (1981).

[25] Lowrance, J.D. and T.D. Garvey, in International Conference on Cybernetics and Society, Proc., 6 (1982).

[26] Lowrance, J. D, Dependency-Graph Models of Evidential Support, Ph.D. dissertation, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts, (1982).

[27] Lu, S. Y., H. E. Stephanou, in National Conference on Artificial Intelligence, Proc. AAAI, 216 (1984).

[28] Nagao, Makoto and Takashi Matsuyama, A Structural Analysis of Complex Aerial Photographs, Plenum Press, New York, (1980).

[29] Nozif, A.M. and M. D. Levine, in Pattern Analysis and Machine Intelligence, Trans. PAMI-6(6), 285 (1984).

[30] Matsuyama, T. et. al., in Workshop on Computer Vision of IPS, Proc. (in Japanese) WGCV 36(3), (1985).

[31] Matsuyama, T. and V. Hwang, in Ninth International Joint Conference on Artificial Intelligence, Proc., 908, (1985).

[32] Pearl, Judea, How To Do With Probabilities What People Say You Can't, Tech. Note Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles, California 90024, September (1985).

[33] Selfridge, P. G., Reasoning About Success and Failure in Aerial Image Understanding, Ph.D. dissertation, TR 103, Computer Science Department, University of Rochester, September (1982).

[34] Shafer, G., A Mathematical Theory of Evidence, Princeton University Press, Princeton, New Jersey, (1976).

[35] Shafer, G., Philosophy of Science, PSA 2, (1978).

[36] Shortliffe, E. H. Computer-Based Medical Consultations: MYCIN, Elsevier Scientific Publishing Co., New York, (1976).

[37] Sloan Jr., Kenneth Robert, World Model Driven Recognition of Natural Scenes, Ph.D. dissertation, The Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, (1977)

[38] Strat, Thomas M., in National Conference on Artificial Intelligence, Proc. AAAI, 308 (1984).

[39] Tanimoto, Steve L., Computer Vision Systems, Edited by Hanson, A. R. and E. M. Riseman, Academic Press, New York, (1978).

[40] Waltz, David L., Computer Vision Systems, Edited by Hanson, A. R. and E. M. Riseman, Academic Press, New York, (1978).

[41] Wesley, L.P., in Workshop on Computer Vision: Representation and Control, Proc., 14 (1982).

[42] Wesley, L.P., in Eighth International Joint Conference on Artificial Intelligence, Proc. IJCAI, 203 (1983).

[43] Wesley, L.P., Lowrance, John D., and Thomas D. Garvey, Reasoning About Control: An Evidential Approach, SRI Technical Note # 324, Artificial Intelligence Center, SRI International, Menlo Park, California 94025, (1984)

[44] Wesley, L.P., Evidential-Based Reasoning in Knowledge-Based Systems, Ph.D. dissertation in preparation, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, (1985)

[45] Weymouth, T.E., Griffith, J.S., et. al., in National Conference on Artificial Intelligence, Proc. AAAI, 429 (1983).

[46] Yakimovsky, Y. and J.A. Feldman, in Third International Joint Conference on Artificial Intelligence, Proc. IJCAI, 580 (1973).

[47] Zadeh, L.A., Information Control, 8, 338 (1965).

[48] Zucker, S.W., Pattern Recognition and Artificial Intelligence, Edited by C.H. Chen, Academic Press, New York, (1977).