To the Graduate Council:

I am submitting herewith a dissertation written by Muhamad B. Abdulghafour entitled "Data Fusion Through Fuzzy Reasoning Applied to Feature Extraction from Multi-sensory Images." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical and Computer Engineering.

M. A. Abidi, Major Professor

We have read this dissertation and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor
and Dean of The Graduate School

DATA FUSION THROUGH FUZZY REASONING APPLIED TO FEATURE EXTRACTION FROM MULTI–SENSORY IMAGES

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Muhamad B. Abdulghafour

December 1992

To my uncle:

The late Mr. Ahmed Abdulghafour

Acknowledgments

This project was partially supported by the DOE's University Program in Robotics for Advanced Reactors (Universities of Florida, Michigan, Tennessee, Texas, and the Oak Ridge National Laboratory) under grant No. DOE DE-FG02-86NE37968.

The completion of this dissertation would not have been possible without the input and responsiveness of many professors and colleagues at the University of Tennessee, Knoxville (UTK). Particular thanks are due to professor M. A. Abidi, who supervised my dissertation and was a constant source of support and advice. I must also express my thanks to professors J. M. Bailey, D. W. Bouldin, Y. Kuo, and M. Trivedi. They served on my doctoral committee and offered to read, critique, and revise this dissertation.

I would like to thank many of my colleagues for their help and encouragement, particularly Mohammed Baccar, Hal Beck, Tan Chandra, ChuXin Chen, Mostafa Rahbar, Hamed Sari–Sarraf, Mukul Shirvaikar, John Sluder, and Mourad Takla.

I am deeply grateful to my wife Nabawie Ibrahim–Basha for her understanding, patience, support, and love.

I will be responsible for any errors or weaknesses that might be discovered hereafter.

Abstract

Multi–sensor systems provide a purposeful description of the environment that a single sensor cannot offer. Fusing several types of data enhances the recognition capability of a robotic system and yields more meaningful information otherwise unavailable or difficult to acquire by a single sensory modality. Because observations provided by sensors are uncertain, incomplete, and/or imprecise, we adopted the use of fuzzy sets theory as a general framework to combine uncertain measurements. We developed a fusion formula based on the measure of fuzziness. The fusion formula was mathematically tested against several desirable properties of fusion operators. We established a fuzzification scheme by which different types of input data (images) would be modeled. This modeling process was essential in providing suitable predictions and explanations of a set of observations in a given environment. A defuzzification scheme was carried out to recover crisp data from the combined fuzzy assessment. This approach was implemented and tested with real range and intensity images acquired by an Odetics Laser Range Scanner. The goal was to obtain better scene descriptions through a segmentation process of both images. Despite the low resolution of the input images and the level of noise associated with the acquisition process, the segmented output picture should be suitable for recognition purposes. Other data fusion approaches such as the Super Bayesian Approach (SBA) and Dempster's rule of combination were implemented and evaluated. A systematic method for evaluating and comparing segmentation results was presented. Various levels of noise were added to the real data and segmentation results from all three approaches were compared and evaluated. The strengths and weaknesses of each method in combining evidence and managing uncertainty were discussed.

TABLE OF CONTENTS

CHAPTER		PAGE
1. Introduction		**1**
1.1	Statement of the Problem	1
1.2	Objectives	4
1.3	Related Work	6
1.3.1	Fusion through Structural Analysis	6
1.3.2	Fusion Using Uncertainty and Imprecision Models	15
2. Data Fusion Through a Test of Fuzzy Hypotheses		**29**
2.1	Fuzziness and Classical Set Theory	30
2.1.1	Preliminaries: Definitions and Concepts	31
2.1.2	Desirable Properties of a Fusion Technique	34
2.2	Data Fusion: A New Fusion Technique	41
2.2.1	Test for the Desirable Properties	44
2.3	The New Fusion Technique vs. Other Methods	51
2.3.1	The Super Bayesian Approach	51
2.3.2	Dempster's Rule of Combination	55
2.4	Fuzzification: Modeling Uncertainty	63
2.4.1	Membership Functions	64
2.4.2	Representation of Fuzzy Values	65
2.5	Fuzzy Reasoning: Data Fusion through Fuzzy Rules Combination	68
2.6	Defuzzification	69

CHAPTER PAGE

3. The Use of Data Fusion for Segmentation: Experimental Results . 73

3.1 Edge Detection 74

3.1.1 The Gradient Operator 74

3.1.2 The Laplacian Operator 75

3.1.3 Zero–crossings (The Marr and Hildreth Approach) 75

3.1.4 Canny 76

3.1.5 Angle of Curvature 77

3.2 Edge Features Classification 82

3.2.1 Jump Edges 82

3.2.2 Roof Edges 86

3.2.3 Texture Edges 86

3.3 Data Fusion System Implementation 91

3.3.1 Fuzzification 91

3.3.2 Fusion 92

3.3.3 Defuzzification 97

3.4 Image Segmentation 97

3.5 Experimental Results Using Other Data Fusion Approaches 99

3.5.1 The Super Bayesian Approach 101

3.5.2 Dempster's Rule of Combination 104

3.6 Experimental Evaluation 106

4. Effects of Fusion Parameters on Segmentation 111

4.1 Effects of Noise Level 111

4.1.1 Experimental Evaluation 112

4.2 Effects of Combination Weighting (α) 116

CHAPTER PAGE

4.2.1 Experimental Evaluation 117

4.3 Effect of the Number of Fuzzy Values (subsets) 119

4.3.1 Experimental Evaluation 122

5. Summary and Conclusions 126

BIBLIOGRAPHY 130

APPENDICES 138

A. Modeling Uncertainty: The Probability Theory 139

A.1 Probabilities: Concepts and Properties 139

A.2 Probability Rules for Generating Compound Events 140

A.3 Joint and Conditional Probability 141

A.4 Independence 142

A.5 Likelihood Functions 143

A.6 Bayes' Theorem 143

A.7 Subjective and Prior Probabilities 145

A.8 Noninformative Priors 147

B. Modeling Uncertainty: The Evidence Theory 150

B.1 Propositions and the Frame of Discernment 150

B.2 Belief Functions 151

B.3 Lower and Upper Probabilities: Credibility and Plausibility 153

B.4 Bayesian as a Special Case of the Theory of Evidence 154

C. Modeling Uncertainty: The Possibility Theory 155

CHAPTER PAGE

C.1 The Fuzzy Sets Theory 155

C.2 Fuzzy Sets Operations 160

C.3 A Possibility Distribution Concept 162

D. Multi–sensory Fusion Using Uncertain Models 168

D.1 Data Fusion Example 169

D.2 Combining Evidence Using the Probability Theory 172

D.3 Combining Evidence Using the Evidence Theory 184

D.4 Combining Evidence Using the Possibility Theory 195

D.5 Averaging Operators 196

E. Comments on Appendices A – D 204

VITA 208

LIST OF TABLES

TABLE		PAGE
2.1	*Dispute measures*	32
2.2	*Values of the derivative for different values of* ϵ	59
2.3	*Summary of combination techniques against the desirable properties*	62
3.1	*Results of segmentation through different data fusion approaches.*	107
4.1	*Results of segmentation after adding noise* ($\sigma = 0.8$).	112
4.2	*Results of segmentation after adding noise* ($\sigma = 1.2$).	115
4.3	*Summary of noise effects on segmentation.*	116
4.4	*Results of segmentation for different valuse of* α.	117
4.5	*Results of segmentation (pixels per region).*	125

LIST OF FIGURES

FIGURE **PAGE**

1.1 *Feature extraction: (a) intensity image; (b) range image; (c) surface normal of range; (d) edges from intensity; (e) edges from range; (f) edges from surface normal* 3

1.2 *A fusion system based on fuzzy reasoning.* 5

2.1 *A modeling example.* 31

2.2 *A new data fusion function* 44

2.3 *A modeling example: (a) three, (b) five, and (c) seven fuzzy values.* . . . 66

2.4 *Fuzzification: (a) video edge map; (b) histogram for (a) plotted as the ordinate to a logarithmic scale; (c) an approximated histogram for (a); (c) fuzzy-set values.* 68

2.5 *Four outputs generated by four rules.* 72

3.1 *The angle of curvature: (a) vectors for computing angles θ_1-θ_4; (b) an illustration in 3-D.* 78

3.2 *Edge detection: (a) intensity image; (b) edges of (a) using the Sobel operator; (c) edges of (a) using the Laplacian operator; (d), (e), and (f) edges of (a) using Canny's edge detector for $\sigma = 0.92, 0.98, 1.2$, respectively; (g), (h), and (i) edges of (a) using Marr's edge detector for $\sigma = 0.6, 0.65, 0.7$, respectively.* 79

FIGURE **PAGE**

3.3 *Edge detection: (a) range image; (b) edges of (a) using the Sobel operator; (c) edges of (a) using the angle of curvature algorithm; (d), (e), and (f) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.93, 0.95$, *respectively; (g), (h), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.67, 0.69, 0.7$, *respectively.* 80

3.4 *Edge detection: (a) normal of a range image; (b) edges of (a) using the Sobel operator; (c) edges of (a) using the angle of curvature algorithm; (d), (e), and (f) edges of (a) using Canny's edge detector for* $\sigma = 0.7, 0.85, 0.9$, *respectively; (g), (h), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.5, 0.6, 0.7$, *respectively.* 81

3.5 *Jump edges: (a) intensity image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.* 83

3.6 *Jump edges: (a) original range image; (b) ideal edge map; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.93, 0.95$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.67, 0.69, 0.7$, *respectively.* 84

3.7 *Jump edges: (a) normal of a range image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.7, 0.85, 0.9$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.5, 0.6, 0.7$, *respectively.* 85

3.8 *Roof edges: (a) original range image; (b) ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.93, 0.95$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.67, 0.69, 0.7$, *respectively.* 87

3.9 *Roof edges: (a) intensity image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the Laplacian operator; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.* 88

3.10 *Roof edges: (a) normal of a range image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.7, 0.85, 0.9$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.5, 0.6, 0.7$, *respectively.* 89

3.11 *Texture edges: (a) intensity image; (b) ideal edge map; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the Laplacian operator; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.* 90

3.12 *Four membership functions for three inputs and one output.* 93

3.13 *A rule base matrix:* θ_1 *and* θ_2 *represent the normal and video edge maps, respectively.* 95

3.14 *Rule base matrices: (a)* $\theta_1 \equiv WE$; *(b)* $\theta_1 \equiv MO$; *(c)* $\theta_1 \equiv ST$. 97

3.15 *Data fusion: (a) edges from video; (b) edges from range; (c) edges from surface normal; (d) Case (1): fused edges of (a) and (c); (e) Case (2): fused edges of (a), (b), and (c).* 98

3.16 *Image segmentation (no data fusion): (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d), (e), and (f) segmented images of (a), (b), and (c), respectively.* 100

3.17 *Image segmentation (after data fusion): (a) edges from intensity; (b) edges from surface normal; (c) fused edges from (a) and (b); (d) segmented image of (c).* 100

3.18 *Image segmentation (after data fusion): (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d) fused edges from (a), (b), and (c); (e) segmented image of (d).* 101

FIGURE **PAGE**

3.19 *Data fusion: (a) intensity; (b) range; (c) surface normal; (d), (e), and (f) edges from (a), (b), and (c), respectively; (g) fused edges from (d) and (f); (h) segmented image of (g); (i) fused edges from (d), (e), and (f); (j) segmented image of (i).* 102

3.20 *Data fusion: (a) intensity; (b) range; (c) surface normal; (d), (e), and (f) edges from (a), (b), and (c), respectively; (g) fused edges from (d) and (f); (h) segmented image of (g); (i) fused edges from (d), (e), and (f); (j) segmented image of (i).* 103

3.21 *Image segmentation using the SBA: (a) edges from surface normal; (b) edges from intensity; (c) fused edges from (a) and (b); (d) the segmented image of (c) (11 extracted region).* 105

3.22 *Image segmentation using Dempster's rule of combination: (a) edges from surface normal; (b) edges from intensity; (c) Fused edges from (a) and (b); (d) the segmented image of (c) (16 extracted regions).* 105

3.23 *Image segmentation: (a) ideal case; (b) SBA; (c) Dempster's rule; (d) the fusion formula.* 109

3.24 *Data fusion: (a), (b), and (c) regions to be fused; (d) using the fusion formula; (e) using the SBA; (f) using Dempster's rule.* 109

3.25 *Data fusion: (a) and (b) regions to be fused; (c) using the fusion formula; (d) using the SBA; (e) using Dempster's rule.* 110

FIGURE **PAGE**

4.1 *Data fusion: (a), (b), and (c) edges from intensity, range, and surface normal with noise added ($\sigma = 0.8$), respectively; (d) ideal segmentation; (e) fused edges using the SBA; (f) segmented image of (e); (g) fused edges using Dempster's rule; (h) segmented image of (g); (i) fused edges using the fusion formula; (j) segmented image of (i).* 113

4.2 *Data fusion: (a), (b), and (c) edges from intensity, range, and surface normal with noise added ($\sigma = 1.2$), respectively; (d) ideal segmentation; (e) fused edges using the SBA; (f) segmented image of (e); (g) fused edges using Dempster's rule; (h) segmented image of (g); (i) fused edges using the fusion formula; (j) segmented image of (i).* 114

4.3 *Effects of α: (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d) fused edges ($\alpha = 1$); (e) fused edges ($\alpha = 5$); (f) fused edges ($\alpha = 7$); (g) segmented image of the ideal case (h) segmented image of (d); (i) segmented image of (e); (j) segmented image of (f); (k) fused edges ($\alpha = 3$); (l) segmented image of (k).* 118

4.4 *Membership functions for three inputs and one output.* 120

4.5 *Rule base matrices: columns and rows indexed by the values of θ_1 and θ_2, respectively; (a) $\theta_3 \equiv VW$; (b) $\theta_3 \equiv WE$; (c) $\theta_3 \equiv MO$; (d) $\theta_3 \equiv ST$; (e) $\theta_3 \equiv VS$.* 123

4.6 *Image segmentation (after data fusion using five fuzzy values): (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d) fused edges from (a), (b), and (c); (e) segmented image of (d) (10 extracted regions).* 124

D.1 *An uncertain model for the unknown θ.* 172

FIGURE **PAGE**

D.2 *(a) judgmental models in relative agreement; (b) judgmental models in disagreement.* 173

D.3 *(a) L.O.P. method for experiment $\mathcal{A}$; (b) L.O.P. method for experiment $\mathcal{B}$.* 176

D.4 *(a) I.O.P. method for experiment $\mathcal{A}$; (b) I.O.P. method for experiment $\mathcal{B}$.* 178

D.5 *(a) H.O.P. method for experiment $\mathcal{A}$; (b) H.O.P. method for experiment $\mathcal{B}$.* 179

D.6 *(a) Super Bayesian Approach for experiment $\mathcal{A}$; (b) Super Bayesian Approach for experiment $\mathcal{B}$.* 184

D.7 *Probability assignment for both knowledge sources.* 185

D.8 *Composition of probability from both knowledge sources (Dempster's rule of combination).* 187

D.9 *Composition of probability from both knowledge sources committed to the same event (Bernoulli's rule of combination).* 190

D.10 *Composition probability of heterogeneous evidence for the case of $A \cap B \neq \emptyset$.* 191

D.11 *Composition probability for conflicting evidence ($A \cap B = \emptyset$).* 192

D.12 *(a) Dempster's rule of combination for experiment $\mathcal{A}$; (b) Dempster's rule of combination for experiment $\mathcal{B}$.* 194

D.13 *(a) the symmetric sum: min operator experiment $\mathcal{A}$; (b) the symmetric sum: min operator for experiment $\mathcal{B}$.* 202

D.14 *(a) the symmetric sum: max operator experiment $\mathcal{A}$; (b) the symmetric sum: max operator for experiment $\mathcal{B}$.* 202

D.15 *(a) combination criteria from fuzzy premises for experiment $\mathcal{A}$; (b) combination criteria from fuzzy premises for experiment $\mathcal{B}$.* 203

CHAPTER 1

Introduction

1.1 Statement of the Problem

Image segmentation is the process of dividing an image into regions, each of which corresponds to a "homogeneous" surface in the scene. Reliable image segmentation is essential for scene understanding. In general, scene understanding is a difficult task because of noise and resolution problems associated with sensors, lack of geometric features of some objects in the scene due to projection of the sensed data from higher dimensionality to lower dimensionality spaces, illumination and reflection for certain sensors, and similarities among objects in the same scene. Intensity images acquired by a video camera are not able to provide three–dimensional information about a scene. Likewise, range images acquired by a laser range scanner are not able to reveal information about shadows and intensity variations in the scene. Using a multi–sensory system to integrate several types of data should enhance the recognition capability of a robotic system and yield more meaningful information otherwise unavailable or difficult to acquire by a single sensory modality. Since sensory measurements are uncertain, it would be more suitable to deal with the problems of managing uncertainty and combining measurements through non–deterministic approaches. The central problem in managing uncertainty is how to combine evidence using incomplete information. Existing techniques have difficulties coping with many real world fusion problems. More specifically, there are common problems such as resolving conflicts among pieces of ev-

idence that is done on the basis of averaging without testing for certainty. This may not be a suitable procedure for the integration of data with extreme dispute. Other techniques have similar problems such as dependency of information when using Bayes' theory or Dempster's rule of combination. Aggregation of assessments in the Bayesian sense does not have obvious symmetry in the treatment of likelihoods and priors. Representing "ignorance," failing to distinguish between "disbelief" and "lack of belief," and conditioning likelihoods with respect to priors are some of the problems associated with Bayes' rule of combination. Unlike Bayesian theory, the theory of evidence (belief) does not allow *a priori* knowledge to dominate the inference process. The theory of belief permits nonadditive degrees of belief and emphasizes the combination of evidence instead of conditioning. Distinguishing between the disbelief and the lack of belief and allowing for conditioning priors in the light of a new proposition contribute to the novelty of Dempster's rule of combination. However, using the theory of belief (through the orthogonal sum), the combined assessment is often an overestimated one. Also, the consensus evidence strength is very sensitive to the numerical values of the inputs; *i.e.*, a small change of the input values may cause a drastic change in the output values. In addition, Dempster's rule of combination does not test for the reliability of information sources. Other approaches are available in the fuzzy sets theory. Although fuzzy sets approaches seem to have more flexibility than those in probability and evidence theories, their performances for fusing contradictory information are by no means completely satisfactory. However, one could utilize a "hybrid technique" consisting of at least two types of aggregating fuzzy operators for data fusion.

Scene understanding requires the extraction of reliable features from the scene. For illustration purposes, we have acquired several images using an Odetics Laser Range

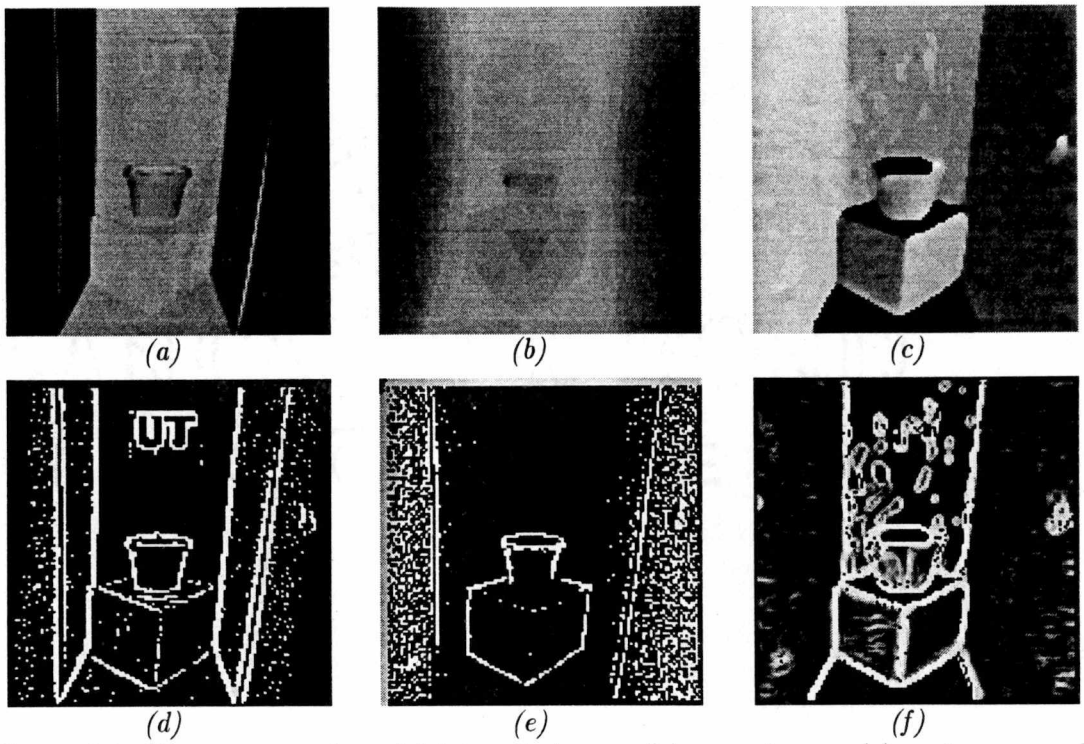


Figure 1.1: *Feature extraction: (a) intensity image; (b) range image; (c) surface normal of range; (d) edges from intensity; (e) edges from range; (f) edges from surface normal*

Scanner. Without considering data fusion algorithms, we have examined the capabilities of single or separate images to provide a "complete" feature map. Range and video images along with a surface normal image are shown in Fig. 1.1. The surface normal image is computed from the range image. Feature edge maps are extracted using Canny's and sobel operators. Glancing through the results illustrated in Fig. 1.1, we observe that each and every edge map has failed to reveal at least one important piece of information about the scene. The edge maps from the range and the surface normal images report no information about texture features in the scene. Likewise, the edge map from the video image has failed to provide complete information about roof edges. The lack of completeness associated with the use of separate sensory modalities

demonstrates the necessity of combining (fusing) several feature maps to generate more reliable information about the scene.

1.2 Objectives

The objective of this dissertation is to develop an improved technique for image segmentation using multi–sensory images. Feature maps from different images can be combined to reduce the uncertainty and vagueness that might be associated with single or separate sensors. After segmentation, homogeneous regions are isolated and labeled. We will derive a new approach for data fusion. This method is based on a test of degrees of certainty associated with measurements. We will show that our approach has such a clear advantage that no ideal assumptions about fusion are needed. We will also prove that our fusion formula possesses several desirable properties that other approaches do not have. Specifically, we will show that the fusion formula is flexible, general, convex, robust, symmetric, idempotent, monotonic, and a regular quazi–arithmetic mean. These mathematical properties ensure a consistent and stable data fusion procedure.

Range and intensity images with exact registration are acquired by an Odetics Laser Range Scanner. Exact registration of data is very important because it guarantees the same appearance, scale, and content of various types of images of the same scene. Feature maps such as roof, jump, and texture edges are extracted. Raw data can be represented by fuzzy models. This modeling process is appropriate to solve the aggregation problem. In these models, each pixel value in the feature maps is assigned a fuzzy membership value within the unit interval $[0, 1]$ (*i.e.*, "fuzzification"). To recover crisp data from the output fuzzy models, we develop a procedure that is able to remap

the output fuzzy model of each feature to a feature space (*i.e.*, "defuzzification"). The latter is a reverse process to the procedure that is used to generate the fuzzy models. The output of this multi-sensor system is a combined feature map for all given inputs. The proposed fusion system is illustrated in Fig. 1.2. The combined map is used as

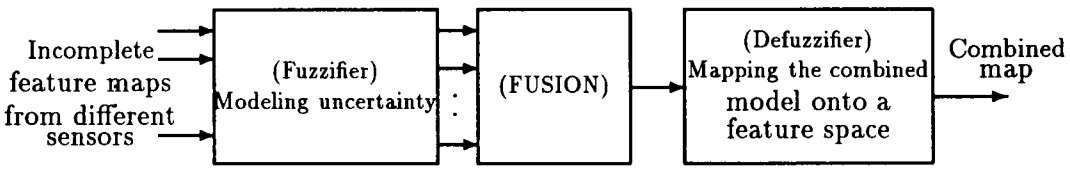


Figure 1.2: *A fusion system based on fuzzy reasoning.*

a reliable input for a segmentation procedure. In the latter, the scene is divided into several homogeneous regions each of which is isolated and labeled. This output image should have enough discriminatory features to simplify a recognition process. To show the effectiveness of this multi-sensory system, the segmentation procedure is applied to each input feature map separately. Results from both procedures (*i.e.*, with and without the data fusion system) are evaluated and compared. Other approaches for data fusion, such as the Super Bayesian Approach (SBA) and Dempster's rule of combination, will be implemented. We will develop a systematic method based on some statistical reasoning in order to evaluate and compare results obtained from our approach and those obtained from the SBA and Dempster's rule of combination. The experimental evaluation is carried out by statistically comparing the approaches' actual results with ideal segmentation results obtained synthetically. Furthermore, we will study the effect of noise on the fused images. The strengths and weaknesses of each method in combining evidence and managing uncertainty were discussed.

1.3 Related Work

In recent years, the problem of data fusion has generated an increasing interest by many researchers. It became a major focus for those who design robot–controlled environments in particular. Having a general solution to data fusion adds considerable intelligence and a functional sophistication to machines working in these environments. The data fusion problem has been defined and treated in various ways by different researchers. For instance, some researchers have used multi–sensory data to help divide a segmentation problem into subtasks. The segmentation was subsequently carried out using a specific type of data that is suitable for a specific subproblem. Other researchers have considered multi–sensory data fusion as a technique that combines data for increasing confidence factors in a recognition process. There is mutual consensus among researchers dealing with data fusion that the ultimate goal is a more efficient recognition system. Approaches toward that goal have varied. The related work in data fusion can be classified into two categories: (1) fusion through structural analysis and (2) fusion through uncertainty and imprecision models. In the following, we review the major papers which deal with data fusion through different approaches. These papers are arranged chronologically.

1.3.1 Fusion through Structural Analysis

The data fusion problem was first addressed through various structural analysis techniques, which were purely deterministic. In the following, we briefly discuss some of the efforts towards this end.

I. Fusion Using Hierarchical Procedures

Garvey [30] described a procedure for locating objects using both reflectance and range data. Two procedures are followed: First, a planning strategy is defined; it includes the selection of features that are likely to belong to the desired object, eliminating features that are common between the desired object and other objects in the scene, and outlining the border of the desired object. Second, all pieces of information are gathered, and the desired plan found by the first strategy is executed. This approach is based on a hierarchical procedure of planning and execution. It creates a graph whose nodes are program modules that gain the desired information at each step. The nodes in the planning graph are called executable subgoal modules. As the planning graph flows, a scoring routine is proposed to determine which branch is best to follow. Detectors that measure some attributes and determine their validity within some desired limits are introduced. An example is given that shows how this technique of combining information in a multi–sensory environment can be carried out to detect and locate an object. The objective here is to establish some relations among objects in the scene that are stored for comparison purposes. A black–and–white picture is taken in exact registration with a range image showing the (x, y, z) coordinates for each point in the image, providing the height z at each point. The object in question has a certain height, which is different from the height of other objects. Information about the area that is provided by the reflectance data is used to detect the desired object. By eliminating the undesired objects, based on the area measure, the author uses a region growing technique to outline the target object.

Gil *et al.* [35] presented experimental procedures to combine intensity and range edge maps. Their approach can be summarized as follows. First, edge maps from

both types of data are extracted. Second, edges from both types of data are fused to enhance the appearance of edges. In detecting intensity edges, the Kirsch operator is applied on a 3×3 neighborhood to calculate the gradient at each point in the image. They calculate the direction of the largest magnitude of the gradient to determine the orientation of each point. Following this process, a thresholding technique is used to eliminate undesirable noise. The method of non–maxima suppression is used to confirm the existence of edge points. There are two types of range edges. The first type of range edge may occur at the intersection of two planar surfaces. This type of range edge is extracted by computing what is referred to as the angle of curvature, which is computed at every point in the range image. A maximum value of the angle of curvature in a neighborhood indicates the existence of an edge. The second type of range edge is what is referred to as the "jump boundary" edge. The extraction of this type of edge is done in two steps. First, sudden changes in range points are assigned the maximal value of the angle of curvature (*i.e.*, 180°), indicating the presence of a jump boundary. Second, a semi–thresholding technique is applied to enhance the appearance of the resulting edges. In implementing this technique, the output range edge map picture will have very thick edges that need to be thinned. For extracting common edges between the intensity and range edge maps, two approaches are used. In the first approach, they utilize a local AND operator by which a pixel from the intensity edge map is ANDed with a window of size $K \times K$ from the range edge map. If there exists an edge pixel inside the $K \times K$ window in the range edge map, then the existence of an edge pixel is confirmed and written as an edge pixel in the output picture. In the second approach, the authors utilize a global AND operator, which is applied to both types of images by tracing edges in both types of data. The tracing algorithm generates a coordinate list

of edge points in the intensity edge map that needs to be confirmed with edge points in the range edge map. This method helps to fill gaps in the edges. The final output image contains all confirmed edges from both types of data.

Magee *et al.* [59] described a technique that combines intensity and range images for three–dimensional object reconstruction. An object is compared with a graph model of the object previously stored. This method can be summarized as follows. A graph representation was derived by extracting polyhedral vertices as nodes and letting common edge lengths be arc values. Circles are extracted using the Hough transform. For circular features, a three–dimensional Hough space is defined to determine the parameters of possible circular curves in the intensity image. These circles are characterized by the center, radius, and plane of the circle. For linear features, the Kirsch gradient is applied to determine edges. Then, the Hough transform is applied to find colinear points which can be fitted into line segments. Once all line segments are found, the intersections among them are found to determine nodes for the graph model. The authors extract three points in the range image that correspond to three edge pixels in the intensity image. These three points in the range image are used to fit a circle in a three–dimensional space. At this point and in a graph containing only circles, partial matching of circles of similar radii is done. For graph matching and model selection, they employ a certain measure for each node in the graph. Thus, they use a two–dimensional intensity image to guide in the detection of important features in the range data. The resulting picture is a reconstructed graph for the scene which can be compared to a stored graph model.

Allen [4] proposed a strategy to organize sensors and sensory–related tasks as classes and methods in an object–oriented programming system. The goal is to avoid changing the system design as the task changes. This would yield an expandable framework

to handle additional tasks. He also describes a framework for implementing a multi-sensory system in two proposed logical high level virtual sensors: an intelligent hand for grasping and a mobile eye that is used for motion tracking. The intelligent hand utilizes a tactile sensor and "finger servos" to control coordinating grasping and tactile recognition. The mobile eye utilizes a vision camera. The main idea of Allen's discussion is to relate sensing to actuation by treating these entities as objects in an object-oriented programming system. His ideas of object-oriented systems is a set of entities (objects) that communicate arbitrarily via messages that allow object methods to be invoked depending on the message content. Several level classes that utilize more than one sensor are defined. For instance, lower level sensor classes seek to find a point of contact based on texture, focus, or other vision operators. Another level seeks to find edges using both vision and tactile sensors. Other sensor classes identify surfaces, which can be accomplished using either sensor. In addition, if a new device is very different from the working devices, a new sensor class or possibly new methods can be added to the system. The system can be expanded for new tasks and new sensors. An example of locating and grasping disks was given. First, stereo vision is employed to locate the three-dimensional position of disks. Using tactile sensors mounted on the robot hand, the robot is able to finely locate and grasp the disks. The tactile sensor would compensate for errors that could occur due to the vision system inaccuracies.

II. Fusion through Complementary Procedures

Duda and Nitzan [23] presented three experiments of using range and intensity data for the purpose of low-level segmentation. First, jump boundaries can be extracted by detecting large changes in the range data which, will reveal only the occluding boundary of large changes in the range and exclude the interior or exterior edges of small changes.

In the second experiment, normal views are extracted by fitting the three–dimensional data to a plane using range information. The reflectance of each point subsequently is computed using the corresponding intensity array. In the third experiment, horizontal planes are extracted by detecting regions whose points have constant heights in the range data. The paper presents a sequential procedure for processing registered range and intensity data to extract regions that correspond to planar surfaces.

Duda *et al.* [24] presented a method to measure and store both intensity and range data. Using a modulated laser beam, the phase shift between the transmitted and received reflected light is measured in order to obtain the depth of each point in the scene. Furthermore, by measuring the amplitude of the received reflected light, they are able to obtain the intensity data in exact registration with the range data. In this particular intensity image, we note the absence of shadows due to the use of the laser range finder itself to compute the amplitude of the reflected light. As a result, variation due to shadows is not detectable. After a simple data transformation in both images, the authors obtained a set of three–dimensional points that represent the range data and a corresponding set of two–dimensional points that represent the intensity data. The paper also describes a sequential procedure to detect and fit planar surfaces. The method starts with a suitable histogram for range data to detect major horizontal planar surfaces according to their Z distribution. Following the detection of major horizontal planar surfaces, a procedure of several steps to fit and refine those surfaces takes place. The authors also use the Hough transform to detect major vertical surfaces. The same procedure of fitting and refining is repeated. The authors avoid using range data to detect slanted surfaces and small regions because the process is computationally expensive. The fitting and refining processes, in the case of intensity data, are done

using a least–mean–square method for all connected components. The paper describes a complete structural method for using both range and intensity data to detect, fit, and refine planar surfaces. The eventual result is a partitioned scene into regions that correspond to planar surfaces of different orientations.

Aggarwal and Magee [3] describe a method that uses both intensity and range images in order to solve the motion problem. Their method is based on constructing a three–dimensional graph, which in turn will be matched with several object models that have already been stored in a data base. The authors use the intensity data to detect all important features that can be used in guiding and extracting three–dimensional information from range data. After finding the three–dimensional information, a graph is constructed. This graph is based on some basic features, such as arc–lengths, angles, and consistency in the arrangement of features. After the three–dimensional graph is constructed, the best match model is chosen from several models of objects, which are already stored to serve as a data base. Once the identification process is determined, the authors calculate the motion parameters using a transformation matrix composed of both translation and rotation parameters. Few assumptions about the characteristics of the motion are made. For instance, the rotation is only around the z–axis. A table of measured and actual readings is constructed for a sequence of images that are taken of a moving object. Using a least squares fit on a larger number of points, the accuracy of the transformation matrix is improved.

Flynn [28] proposed a combination of sonar and infrared sensors in a robot–controlled environment. The technique makes use of the advantages of each sensor thereby utilizing them in a complementary fashion. For instance, a sonar sensor has a low resolution but is a useful tool to measure depth. An infrared sensor, on the other hand, has a

good resolution for detecting objects but is unable to measure distance accurately. A refined map is created by combining the two pieces of data. This map is converted into an intermediate representation such as the curvature primal sketch. The sketch creates a global map for a room by merging views taken from different perspectives. This mapping provides a robot with an intelligent path planner in unstructured environments.

III. Fusion through Other Procedures

Bajcsy and Allen [6] pointed out the important issues needed in developing a multi–sensory system. These issues are representation of object models, organization of the database of models, accessing the database, strategies for using sensors, and convergence of sensory data. In this paper, an example illustrating ways of handling such problems is presented. The focus is on building a rich model description and developing control strategies to allow the sensed data to be matched against a pre–existing model. Multiple access paths into the database can be used to measure convergence and determine degrees of discrimination that have occurred among competing models.

Harmon *et al.* [42] described a mechanism through which data fusion can be accomplished. The authors outlined different cases of fusion and proposed different methods to be suitable and effective. The authors show a general representation for an environment where a robot can be used. The representation gives a general mathematical formulation for objects, errors, measurements, and associated confidence factors. The decision and guidance of the robot are based on the confidence measures, which have been proposed as a primary means for fusing data about the same object properties from different sensors. An intelligent communication interface is proposed to reduce complexity in robot design. In this paper, a distributed blackboard framework is implemented onboard the USMC Ground Surveillance Robot space(GSR).

Bixler and Miller [8] proposed a strategy for using two sensors to guide an autonomous robot in an unstructured environment. First, obstacles are detected using a vision camera. Second, depth between the obstacles and the robot are measured using an ultrasonic range finder. Depth information can gain some clues about the shape of the obstacles. This information is used by a robot to navigate.

Wang and Aggarwal [79] introduced a new algorithm for modeling three–dimensional objects using multi–sensory data. The modeling process is performed in two major steps. First, it extracts three–dimensional surface descriptions from a set of single views. Seconde, it integrates all pieces of information gathered from multiple viewpoints to create a three–dimensional model for the object. At each location where a picture was taken, the occluding contour, which is the projective range of the invisible surface structure, is found. An initial estimate for the shape of the object can be found from the intersections of other occluding contours. To refine these results, they use some constraints on partial structures as the object is viewed from certain directions. Using this method, they are able to determine the position of surface structures in space by comparing the projective range of points from different viewing directions.

Shekhar *et al.* [70] presented a method for the estimation of position and orientation parameters of an object. Their procedure in estimating the parameters consists of two steps. First, intermediate parameters are found by combining measurements. Second, a system of equations is used on the intermediate parameters to obtain the final parameters. A selection of the best intermediate parameters has a considerable impact on obtaining the position and orientation of an object of a known size. The authors assume that correspondence between sensor measurements and object dimensions is given; then the best intermediate parameters are obtained. The least expected error of

estimation has also been determined. This, in turn, is used to define the optimality of this approach.

1.3.2 Fusion Using Uncertainty and Imprecision Models

In this section, a literature survey of data fusion through uncertainty and imprecision models is presented by introducing recent developments in the area. We categorize data fusion approaches according to the theoretical frameworks they use.

I. Fusion through Probabilistic Models

Bolles *et al.* [9] suggested that the problem of segmentation, labeling, and recognition can be decomposed into sub–problems, each of which is an estimation and recognition problem that uses only a portion of the available data. This decomposition allows a parallel computation solution in real time. All operations are combined to achieve globally optimal estimation and recognition. Errors associated with the estimation and recognition of subregions are determined. Objects are approximated by a small number of patches of planes, cylinders, and spheres. Because smoothly varying surfaces can be locally approximated by quadrics, complex objects are composed of these simple 3–D surfaces. Estimated parameters are found for all surfaces in separate parallel processes. The authors also introduced probabilistic models for the generation of range data in order to develop statistical position estimates for 3–D objects. They find the likelihood that is maximum for a certain position. If the right data are taken from a surface in a certain position, then the likelihood is maximum for these data. The authors avoid, however, finding all possible likelihoods for all possible positions and selecting the maximum likelihood due to the high computation cost. The authors suggested that an approximate maximum likelihood estimate is a product of the conditional likelihoods,

which is computationally more feasible. A Bayesian decision is made for each subregion to generate a primitive estimator–recognizer. Pairs of primitives are related by matching to a stored model. For each primitive, there are estimated parameters and associated matrices, derived using matching and combination.

Sher [71, 72] presented a method to compute the probability of edges. Once a likelihood probability is determined, Bayes' rule is used to detect the feature in question. The author demonstrated that the likelihood probabilities generated from different models for the same scene can be combined linearly using a weighted average. These techniques are discussed in more detail in Raiffa [65] and Berger [7].

Luo *et al.* [58] suggested a general technique of using multiple visual and non–visual sensors. They present some basic concepts for acquiring information from the scene through several distinct phases. First, global information about the scene may be acquired from far away. Based on this global information, certain actions are initiated. Seconde, if the acquired global information is not informative enough, then one should "zoom" in and take a closer view of the area of interest to obtain local information. Again, if more information is needed, one must proceed to the next sensing phase, *i.e.*, using another type of sensor. They pointed out that every type of sensor that is used to explore the scene reports both common pieces of data and phase–specific pieces of data. Common observations such as position and orientation might be obtained by different sensors. Phase–specific observation, as its name implies, is only available from one type of sensor. Consequently, they proposed distinct templates, each of which represents a frame type scheme of data for each type of sensor. A basic computational data flow and its architecture are proposed. This architecture includes communication between templates and the fusion process which is used to update the templates and the task

environment as well. In their suggestion for the fusion process, they propose a method for obtaining a measure of confidence associated with each sensor's measurement. The authors suggest fusing data by giving more credibility to sensors that have a higher degree of confidence. Finally, directed graphs are obtained by representing all the sensors used as nodes, then drawing an arrow from one node to another if a consensus between the two nodes is reached. This graph determines a directed graph representation for all sensors that have a consensus about a certain measurement. Thus, the largest connected graph that contains several sensors is selected for use. All sensors (nodes) of this largest connected graph are the best candidates for consultation regarding the data in question. The fused values are used to correct erroneous assessments that are reported by erroneous sensors.

Mann [60] addresses methods for solving multi–sensory integration (MSI) problems. The methods basically intend to solve conflicting assessments that are made through consistent labeling of object features. Bayes' theorem is proposed to maximize the probability of a label given an object feature. This maximum *a posteriori* expresses the confidence level of uncertainty. The author suggested implementing the relaxation process in a parallel computing environment. The integration of labeling between two sensor domains is solved by comparing labels and solving conflicts. For instance, if two labels describing an object feature are in conflict, the label with the lower *a posteriori* probability is discarded. If there is no conflict, the two labels support one another. Registration of features from different sensors for dynamic environments is assumed. Note that the approach allows for changing prior assessments based on the maximum *a posteriori* probability integrated from other sensors. This labeling algorithm allows for parallel implementation. The author points out some other mathematical approaches

for labeling algorithms, such as stochastic relaxation with simulated annealing. The latter depends heavily on a suitable model for the annealing process.

Durrant–Whyte [26] proposed a general framework for modeling observations and making decisions of a multi–sensory integration system under the assumption of uncertain geometric information. The author described a technique for modeling sensors and their observations. Every sensor in the system is treated as an individual decision maker that can communicate with other sensors as a team member. All sensors in the system are sources of many uncertain pieces of information that need to be fused to form a consensus and produce optimal decision–making. The mechanism of manipulation, coordination, and multi–sensory integration of uncertain measurements is performed through interactions among three components of the sensor models. The first component is the observation model; it describes uncertain measurements given by a sensor. The second component is the dependency model, which describes the dependency of a sensor on other sensors in obtaining measurements. The third component is the state model, which describes how the position and the internal state of the sensor affect the measurement acquisition. The author emphasized the drastic simplifications in dealing with fusion if one chooses to use non–deterministic approaches, such as the probability theory. Many multi–sensory data systems utilize sensors of different types. Different types of data ought to be fused. These geographically different types of data are hard to fuse unless all of them can be transformed to a common domain. He strongly emphasizes the importance of such transformations, so that all different sensors can interact and communicate in a dimensionless framework. This common language is derived by transforming all data to a probabilistic model. The author also points out the necessity of having quantitative models that allow for analyzing the system's performance,

determining estimation ability, and developing sensing strategies to serve the task requirements. In contrast, heuristic rule–based approaches are unable to quantify sensor performance. After all types of data are transformed to a common parametric domain, opinions of sensors are fused using a Bayesian classifier where a sensor is considered as a member of a multi–Bayesian team. Contaminated Gaussian distributions are proposed to represent observation models. These distributions are capable of handling not only information corrupted by noise, but also the representation of systematic errors. Models for state and dependency are generally proposed. Finally, the dependency model is suggested to help dynamic changes of information in competitive, complementary, and cooperative interactions.

Elfes [27] proposed a non–deterministic approach for spatial robot perception and navigation. His ideas are based upon constructing the so–called "Occupancy Grid" (OG). The OG is determined by utilizing the Bayes' theory. Because of the uncertainty of incoming readings from different sensors, the state of each cell in the OG is updated. The author also suggests that the OG can be used to guide an autonomous robot in an unknown environment. Decision–making and navigation of the robot solely depend on the state of the OG. The Super Bayesian Approach (SBA) is used to combine two incoming pieces of information to generate the final assessment [7].

Hager and Mintz [39] presented a fusion technique based on an implementation of finite elements of Bayes' theorem. Given an *a priori* probability density over unknown parameters, a new probability density of sensor observation is computed using Bayes' theorem. Given a finite partition of a distribution, the current *a posteriori* probability measure, and an observation, they suggested using Bayes' theorem to compute the associated probability. The independence between the observation and the finite partition

of the distribution is assumed. The authors also discussed the robustness of such a method when applied to uncertain sensor models. Predictable errors and quality measures are mathematically demonstrated. One primary limitation of this technique is the computation cost. This method lends itself to using parallel computation.

II. Fusion through Evidential Reasoning

Garvey *et al.* [31] described an inference technique based on the combination rules of the theory of belief. They showed how sensors interact in a common domain. An updating technique of received information can be employed. The authors define the quantities of belief functions, such as the support function, plausibility, and uncertainty measure. Numerical examples are presented and further investigation for linking sources of information smoothly and flexibly is suggested.

Shafer [69] has shown that Bayesian rules of combination do not treat priors and propositions fairly. In contrast, belief theory allows fair consideration of all pieces of evidence. The author's argument is supported by a practical example. Identifying a glass by its refractive index is presented using both theorems [67]. The author successfully shows that the Bayesian approach of combining pieces of evidence gives a final judgment in favor of the prior probability distribution. The author showed that Lindley's Paradox [69] seems to interpret the diffuse prior as a representation of strong prior evidence. Bayesian rules of combination rely heavily on the priors taking values from empirical frequency distributions. However, he showed that the theory of belief explicitly avoids such dependencies. In this paper, the justification of non–coherent assessments between the occurrence of an event and its negation is explicitly emphasized.

Lowrance *et al.* [57] added to the theory of belief new evidential relations, each of which is of considerable significance. The first one is the compatibility relation,

where propositional statements can be translated from one frame of discernment to another. The second one is the summation relation, which eliminates extraneous details from the mass distribution. The third one is the "gisting" (among others) relation, which produces the "central" Boolean–valued statement that captures the essence of a mass distribution. Finally, a graphical representation is introduced to improve the accessibility of these techniques.

Chatterjee *et al.* [13] presented a comparison between the Dempster–Shafer theory of belief and the extended Dempster–Shafer theory of belief suggested by Ishisuka [46]. This comparison was carried out through the fusion of visible and thermal infrared images. The authors pointed out that some clustering techniques must be made on the spatial distribution of the sensor response for all input images before the fusion process begins. These preprocessing techniques of images, such as segmentation or clustering, are done on the basis of some global feature–space characteristics and/or indirectly derived from portions of the response curves of a given sensor. In this work, a clustering analysis on both types of images is presented. The resulting image has eight dominant classes, each of which has a different shade of gray. A membership value on the basis of gray level is assigned to every pixel in the clustered classes. Two classes of interest are selected from the visible and thermal classes. Finally, these two classes are fused using both the Dempster–Shafer belief theory and the extended Dempster–Shafer belief theory. The comparison between the two results is carried out by computing the difference between them. This shows how Dempster–Shafer's technique overestimates the fused data and how the extended Dempster–Shafer's technique reduces the dependency problem effect.

III. Fusion Using Fuzzy Models

Jain [47] discussed cases of inexact reasoning that statistical decision theory is unable to handle. These cases are considered as such not because of the attribution of uncertainty, but rather uncertainty due to imprecision. These ill–defined cases are no stranger to humans. Linguistic variables such as very small, very large, and medium are examples of these cases. Humans are capable of making decisions and choosing among alternatives. This type of inference is similar to "if statements" in a programming language, which implies reasons, then makes a decision. If the alternatives are too numerous, humans are no better than the statistical decision theory. As a consequence, the author proposes a mechanism by which a systematic inference procedure takes place. The suggested technique here is the use of the fuzzy theory proposed by author(s) [84] which assigns membership values for all alternatives to represent the relative strength of each ill–defined term. The utility of the system is introduced; it relates the system states to all possible alternatives. Maximum utility and/or maximum grade of membership play key roles in optimizing the decision–making. Numerical examples are implemented on the basis of discrete alternatives. First, fuzzy utilities are determined; then the maximizing subsets of a fuzzy set are deduced. Next, a fuzzy set is formed by combining the information of the maximizing subsets for alternatives and the predetermined fuzzy utility. This is done by the intersection of the two fuzzy quantities which, in turn, yields the grade of membership. Finally, the best alternative is chosen, having the highest grade of membership. The author emphasized the use of both maximum grade of membership and maximum utility associated with various alternatives.

Ishisuka *et al.* [46] presented a reasoning technique that is based on the problem–reduction method. In this paper, problems are decomposed into simpler sub–problems,

making inexact reasoning easier to accomplish. A hierarchical procedure for problem–reduction is proposed. The authors pointed out that for some particular domains, inexact reasoning using Bayesian statistics and/or Dempster–Shafer's belief functions is not adequate or effective. For these reasons, the authors propose the use of Bayesian and/or Dempster–Shafer theorems in the process of reasoning in a fuzzy framework. This allows them to reason about conflicting bodies of evidence using fuzzy mathematics. Decisions are made in subproblems to achieve subgoals. These decisions propagate through certain types of graph nodes such as AND/OR/COMB, to eventually provide a reasonable answer to the overall problem. They also proposed that ANDing and ORing two membership values of two pieces of evidence are equivalent to the interpretation of credibility and possibility of measures in a fuzzy environment, respectively.

Yager [82, 83] discussed approaches to approximate reasoning that are based on the conjunction and projection principles, the law of compositional inference, and the entailment principle. A form of inclusion that is associated with the belief is attributed to the data and production rules. The ability of introducing certainty qualification into an expert system is presented. The reasoning methodology is based on Zadeh's representation of quantifiers. Different types of quantifiers are discussed in a linguistic manner. Certainty qualifications are introduced to each dispositional fuzzy conclusion, which can be transformed into a possibility distribution.

Cohen *et al.* [15] discussed several issues concerning decision–making. One of these issues is the necessity of using fuzzy mathematics as a principal method in cases where traditional mathematics is not applicable. Due to the frequently ill–defined conditions and uncertain behavior associated with inexact reasoning, they argued that a fuzzified version of any statistical concept is a more effective tool in decision–making. Several

concepts in decision analysis are presented: (1) Fuzzy probabilities, such as linguistic quantities, likely, unlikely, and highly improbable, which are imprecisely known or defined. Fuzzy probabilities are used to avoid making imprecise assessments when, for instance, probabilistic numbers are involved. Instead, they allow an analyst to come closer to reality involving decision–making under uncertainty. (2) Fuzzy quantifiers, which are characterized by expressions such as most, few, several, almost all, frequently, and usually. These quantifiers may be interpreted as a relative count of elements in fuzzy or crisp sets. They argue that fuzzy quantifiers give them common sense reasoning that humans use in everyday decision–making. Also, fuzzy mathematics allows them to reason a case by inferring common sense causes. For instance, "icy roads are dangerous" is a dispositional conclusion from dispositional premises that "icy roads are slippery" and "slippery roads are dangerous." (3) Fuzzy truth values, such as quite true, completely false, and somewhat true. In this case, some intermediate truth values can take place. In contrast, propositions in classical mathematics can only be true or false. This wider range of truth values in fuzzy mathematics enhances the formulation of the decision–making process. (4) Linguistic variables whose values are words or sentences in a natural or synthetic language. For instance, values are linguistic variables if they are expressed as young, not young, not very old, etc. The importance of this type of fuzzy description stems from the fact that it gives an effective assessment in ill–defined cases. Acquiring and using information in cases where ideal assumptions are far from reality are discussed in detail.

Prade [64] made a clear distinction between uncertainty and imprecision. When the truth (or falsity) of a given assessment cannot be definite, the assessment is said to be uncertain. If the state of an assessment is a function of some variable whose value is

not sufficiently determined with respect to a given scale, the assessment is said to be imprecise. According to this definition, a precise assessment may be uncertain, and a completely certain assessment may not be precise. From a statistical point of view, the distinction between uncertainty and imprecision must be pointed out because it determines the modeling strategy where different statistical quantities are encountered. By definition, imprecise measurements are encountered in multiple sensor fusion, such as the fuzzy quantities, *i.e.*, possibility and necessity measures. The concern of modeling observations must include uncertain and imprecise quantities and methods of statistical reasoning that can deal with such quantities.

IV. Fusion through Other Inexact Reasoning Techniques

Whalen [80] discussed several methodologies for handling decision–making processes in both well– and ill–defined cases. These methodologies are discussed in terms of their abilities to handle inexact reasoning in particular problems. A comparative summary of several techniques is presented. The author points out that the principle differences among techniques are in the assumptions they make regarding the information that is available at the time of the decision.

Delcroix and Abidi [18] presented a new approach for the fusion of range and intensity edge maps. Their method is based on the interaction between two constraints. The first one is the principle of existence by which the output value of an edge is maximized, if it is detected. The second one is the principle of confirmability, by which the existence of an edge in one type of data is confirmed if there exists a corresponding edge in the other type of data. These two principles are simultaneously optimized using a fusion function. The fusion function is assumed to be analytic, hence can be represented in a converging Taylor series. The two principles are combined by maximizing a linear com-

bination of the two constraints. The maximization is done using the Euler–Lagrange Calculus of Variation equations. The resulting fused edge maps appear to be smooth and accurate. Furthermore, this technique has the advantage of generality, which can be extended to other types of features.

Ayache and Faugeras [5] proposed a method to construct a three–dimensional description for the environment of a mobile robot using a stereo system. Their method allows dealing with geometric information and the corresponding uncertain information in the following manner. First, the nonlinear relation between an observation and some parameters in the three–dimensional space is linearized. Second, a Kalman filter is utilized to find the best estimate of uncertain measurements. The uncertainty is represented by a set of covariance matrices. The most reliable estimate is based on the measurement that has the least value in these matrices. The resulting description is called the realistic uncertain description of the environment. Third, an estimate of the three–dimensional displacements is obtained using several stereo pairs. This yields a relation between the position of the robot and the uncertain characterization of these displacements. Finally, an updating scheme is performed by matching measurements in different frames with the actual feature in some object in the environment.

Cheng and Kashyap [14] discussed a number of desirable properties of the Dempster–Shafer belief theory and Bayesian statistics when applied to real problems in expert systems. Several assumptions have been made, such as independency between bodies of evidence; commutativity and associativity of the combination function; monotonicity, which means that stronger evidence makes a stronger belief function; and the smoothness of the combination function, which is considered a measure of robustness. The authors also define the property of generalized symmetry, which guarantees that a com-

bination function on a hypothesis can be used as a combination function on its negation. Another property is isomorphism, which guarantees one–to–one mapping between two systems. The authors show that under the isomorphism assumption, the derivative of the combination function is unbounded. They conclude that the Dempster–Shafer theory of belief and Bayesian statistics are often not robust. Under certain proposed constraints on the combination function, they are able to define several fusion functions. Finally, under the assumption of isomorphism, there is inconsistency between continuous and discrete inferences. This is rooted in the difficulty of realizing correspondency in relations between the two domains, which in turn violates the fundamental assumptions such as associativity and commutativity.

Grandjean *et al.* [37] utilized an extended Kalman filter to fuse measurements in the presence of noise. Sensors are used in a complementary fashion. A laser range finder is used to detect a surface, and a stereo vision system is used to compute its extent and location via detecting three–dimensional segments on the surface. This methodology relies on two principles: (1) sensor descriptions include a numeric model necessary for the fusion and the modeling algorithm (2) data is propagated at every level of the perception process. Also, using multiple laser cross–sections helps in modeling the whole scene.

There have been a number of papers discussing the problem of reasoning under uncertainty and presenting worthwhile techniques to overcome difficulties associated with this problem. Some of these papers employ the probability theory; others have been presented as a departure from classical decision theory such as the Dempster–Shafer belief theory and the possibility theory. Selected papers and references dealing with probabilistic models are [7, 10, 12, 20, 25, 29, 33, 34, 38, 40, 44, 49, 53, 55, 56, 62, 65, 66,

81]. Papers and references dealing with the belief theory may be found in [19, 22, 67, 69]. Finally, some useful references which deal with the problem of inexact reasoning through the possibility theory and the fuzzy sets theory are [21, 32, 50, 63, 78, 92, 93]. These include a review and a discussion of many probabilistic and non–probabilistic methods in expert systems and other uses.

CHAPTER 2

Data Fusion Through a Test of Fuzzy Hypotheses

Data gathered by different knowledge sources (sensors) are often uncertain, imprecise, fuzzy, vague, or incomplete. In a multi–sensory system, sensors may report different types of data to describe the same scene. In order to combine or fuse several types of data which may be geometrically different, one needs to transform all data to a common space. This transformation is called "modeling." For each sensory reading, the result of a modeling process is an uncertain hypothesis; its truth or falsity is yet to be determined. Many data fusion systems, based on evidence or observations, are able to generate numerical support for hypotheses. Modeling these hypotheses depends on the framework that is used for fusion. Based on the survey we conducted, we have concluded that fuzzy sets theory has an outstanding ability to model the amount of ambiguity or vagueness associated with sensory data in quantifying meanings that allow for numerical computations. Our data fusion system, therefore, combines fuzzy hypotheses through the degrees of support committed to these hypotheses. Modeling data in a fuzzy set framework is different from using the classical set theory. The latter views the occurence of events in black–white logic. The former is more flexible and allows for gradual transition from one logical concept to another. In the following section, we introduce the basics of fuzzy modeling and outline the differences between fuzziness and traditional set theory.

2.1 Fuzziness and Classical Set Theory

To introduce the difference between fuzziness and classical set theory, let us consider the following experiment. Assume that a feature map is extracted from an image. Suppose that the feature values (propositions) can be characterized using linguistic values such as "weak," "moderate," and "strong." While the classical set theory defines sets with a clear–cut concept, the fuzzy set theory allows us to represent set membership as possibility distribution (see Fig. 2.1). In the classical set theory, a feature either belongs to a set or it does not. In other words, the transition from the concept of "belonging" to the concept of "not belonging" to a set is abrupt. When there is a source of vagueness or uncertainty for the belonging proposition, it is not possible to adequately draw any clear–cut borderline. The solution to this problem is suggested by Zadeh [84]. In Zadeh's approach, each subset in a set (universe of discourse) is assigned a degree of membership. The latter can take in values within the unit interval $[0, 1]$, not only 0 or 1. Figure 2.1 shows graphical representations for modeling data in both fuzzy and conventional sets frameworks. From this representation of membership, a fuzzy set may be defined as follows:

Definition: A fuzzy set, A, in a universe of discourse, $\Theta = \{\theta\}$, is denoted by $A \subset \Theta$ and defined as the following set pairs

$$A = \{(\mu_A(\theta), \theta)\},$$

where $\mu_A : \Theta \rightarrow [0, 1]$ is the membership of A and $\mu_A(\theta)$ is the grade of membership of $\theta \subset \Theta$ in A.

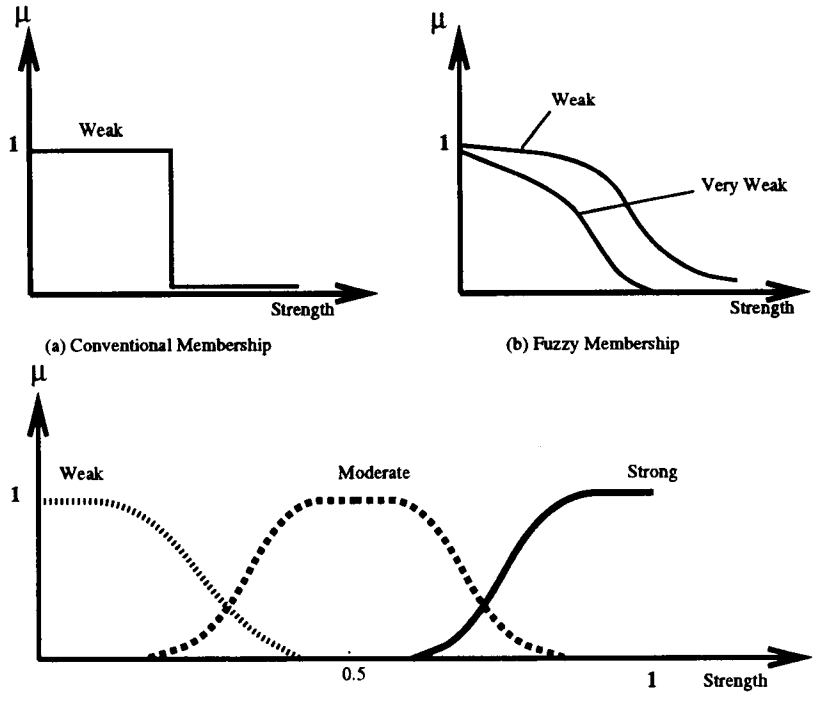


Figure 2.1: *A modeling example.*

2.1.1 Preliminaries: Definitions and Concepts

As a first step to data fusion, we must introduce several definitions and concepts. Not only do these definitions allow us to build a common domain in which approaches applied to data integration can be analyzed, but also they address the problems of managing uncertainty.

1. **Degree of Truth** [67]: $\forall\theta \subset \Theta, i = 1, 2, \ldots, n$, there is a belief function $\mu_i : \Theta \rightarrow [a, b]$ commits a degree of truth to θ inside the truth region $[a, b]$, where n is the number of information sources supporting θ.

2. **Measure of Dispute or Conflict** [67]: The measure of dispute is an indication of how close two or more fuzzy observations are to each other in a given set of observations. The measure of dispute assumes a numerical compatibility among

Table 2.1: *Dispute measures*

Method	Equation
Absolute Distance	$D_{xy} = \|x_1 - y_1\| + \ldots + \|x_n - y_n\|$
Euclidean Distance	$D_{xy} = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$
Minkowski Distance	$D_{xy} = [(x_1 - y_1)^\alpha + \ldots + (x_n - y_n)^\alpha]^{\frac{1}{\alpha}}$
Mhalanobis Distance	$D_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - y_i)(x_j - y_j)$
Functional Distance	$\frac{1}{(1+D_{xy})}$

measurements. To do so, one may choose to compute distances among given observations, which can be determined in several ways. In the literature, there are several methods to measure dispute such as association, correlation, probabilistic, and functional methods. Examples of these methods are illustrated in Table 2.1. In this work, we will adopt the absolute distance, which is defined as follows:

Definition: $\forall\theta \subset \Theta$, the degrees of truth $\mu_i(\theta)$ and $\mu_j(\theta)$, $\forall i, j = 1, 2, \ldots, n$, committed to an event θ, have a degree of dispute $DD_{ij}(\theta)$ given by

$$DD_{ij}(\theta) = |\mu_i(\theta) - \mu_j(\theta)| . \tag{2.1}$$

Definition: Two fuzzy hypotheses, $\mu_i(\theta)$ and $\mu_j(\theta)$, are said to be in severe conflict if and only if the degree of dispute, $DD_{ij}(\theta) > t$, where t is a threshold value which is assigned according to the application domain.

3. **Measure of Fuzziness:** Given a universal set, Θ, a *measure of fuzziness* is defined as

$$f : \mathcal{F}(\Theta) \rightarrow \Re,$$

where $\mathcal{F}(\Theta)$ denotes the set of all fuzzy subsets of Θ. This means that the function f assigns a value $f(\theta)$ to each fuzzy subset A of Θ. For both $\mu(\theta) = 0$ and $\mu(\theta) = 1$, $\mu(\theta)$ is a non–fuzzy value, and therefore, $f(\theta)$ must be minimum. A non–fuzzy value is sometimes called a crisp value. If the truth region is defined within the unit interval $[0, 1]$, the identity value is at $\mu(\theta) = e = 1/2$ which is considered maximally fuzzy. The identity value (the midpoint value) is unique. The fuzzier A is, the closer $\mu_A(\theta)$ is to the midpoint of the fuzzy support. Likewise, the less fuzzy A is, the closer $\mu_A(\theta)$ is to the nearest crisp value (*i.e.*, 0 or 1) [52]. According to this argument, we define the measure of fuzziness $f_i(\theta)$ associated with each fuzzy membership $\mu_i(\theta)$ in terms of its metric distance from the crisp points. That is,

$$f_i(\theta) = 1 - |2(\mu_i(\theta) - e)|, \quad \forall i = 1, 2, \cdots, n\,. \tag{2.2}$$

Other metric distances to measure the degree of fuzziness may be found in [51]. Our definition to the measure of fuzziness is, in essence, compatible with the measure of fuzziness proposed by De Luca and Termini in 1972, which is given by

$$f_i(\theta) = -[\mu_i(\theta)\log_2\mu_i(\theta) + (1 - \mu_i(\theta))\log_2(1 - \mu_i(\theta))]\,. \tag{2.3}$$

The normalized version of f within the unit interval $[0, 1]$ is given by

$$\hat{f} = \frac{f}{|\Theta|},$$

where $|\Theta|$ denotes the cardinality of the universe of discourse Θ. Equation 2.3 is also known as the Shannon's function, which is monotonically increasing in the interval $[0, e]$ and monotonically decreasing in $[e, 1]$ with maximum at $\mu_i = e$. Note that Eq. 2.3 is not defined at the crisp points of the truth region.

2.1.2 Desirable Properties of a Fusion Technique

Many data fusion systems rely on formulas that combine evidence. In this section, several desirable properties, such as robustness, test for certainty, symmetry, convexity, generality, and monotonicity, are discussed. Studying these desirable properties is the main theme for a mathematical analysis and an experimental evaluation of our approach, and other approaches, to data fusion. Let Θ be the sample space which contains a collection of events; *i.e.*, $\Theta = \{\theta\}$. Assume that $\mu_1(\theta), \mu_2(\theta), \cdots, \mu_n(\theta)$ are degrees of truth over Θ committed to an event θ in Θ. The desirable properties are as follows:

Generality: A fusion technique should handle any number of pieces of evidence with high efficiency of computation. The combined evidence $\mu(\theta)$ for n pieces of evidence is given by

$$\mu(\theta) = f\left[\mu_1(\theta), \mu_2(\theta), \ldots, \mu_n(\theta)\right] , \tag{2.4}$$

where f is a combination operator that is able to combine n pieces of evidence simultaneously. Sequential computations often increase the number of operations required to combine data. For instance, using an algorithm that is based on a pairwise computation, one must repeatedly apply the algorithm to combine all pieces of evidence. Furthermore, a combination function should be able to cope with different application domains. For instance, it is sometimes desirable that one applies a certain operation of various strengths.

Test for Certainty: Conflicting and/or non–conflicting pieces of evidence must be fused on the basis of their degrees of certainty. We have already defined that two given assessments, μ_i and μ_j, supporting the same event, θ, have a degree of dispute $DD_{ij}(\theta)$.

In the light of the degree of dispute, we define a distinction between the following two cases:

Non–conflicting Evidence ($DD_{ij} < t$): $\mu_i(\theta)$ and $\theta_j(\theta)$ are said to be in relative agreement if and only if $DD_{ij}(\theta) < t$, where t is an appropriate threshold value. In this case, several methods can be used to fuse data. One method of combining these assessments is to compute the "intersection" of $\mu_i(\theta)$ and $\mu_j(\theta)$ and to normalize the result [22]:

$$\forall\theta \subset \Theta,\ \mu_{ij}(\theta) = \frac{\mu_i(\theta) * \mu_j(\theta)}{\sup_{\theta\in\Theta} \mu_i(\theta) * \mu_j(\theta)}, \tag{2.5}$$

where $*$ is a fuzzy set–theoretic intersection. If $*$ is assigned the min operator, then the combination is based on a "cross–validation" of $\mu_i(\theta)$ and $\mu_j(\theta)$, which are considered to be in agreement. The other method of combination is to compute the mean of these two fuzzy quantities. The mean is defined by mapping m from $[0,1]^2$ into $[0,1]$ such that $\forall\mu_i, \mu_j$

1. $\min(\mu_i, \mu_j) \leq m(\mu_i, \mu_j) \leq \max(\mu_i, \mu_j)$,

2. $m(\mu_i, \mu_j) = m(\mu_j, \mu_i)$

3. m is a non–decreasing function of each of its arguments.

The functional solution for a–c is derived by Aczel [2], and is given by

$$m(\mu_i, \mu_j) = k^{-1}\left(\frac{k(\mu_i) + k(\mu_j)}{2}\right), \tag{2.6}$$

where k is continuous and strictly monotonic. Note that the min and the max operators, arithmetic, geometric, and harmonic means are special cases of Eq. 2.6. In the literature, one can find several approaches to combine non–conflicting evidence, namely,

the Super Bayesian Approach, Dempster's rule of combination, and other techniques in the framework of probability theory.

Conflicting Evidence ($DD_{ij} \geq t$): In real world problems, contradiction among sources of information is not an uncommon occurrence. We believe that if two assessments committed to the same event are in conflict, at least one of these assessments is unreliable. It follows that if conflicting bodies of evidence are given, one must fuse data according to their degrees of certainty. In other words, combinations which are based on ordinary averaging operators may not be suitable. One method proposed by Dubios [22] to solve conflict between pieces of evidence is to apply the fuzzy–set theoretic union. The fuzzy–set theoretic union $\forall\theta \subset \Theta$, is defined by

$$\mu_{ij}(\theta) = \max\left[\mu_i(\theta), \mu_j(\theta)\right] . \tag{2.7}$$

Note that μ_{ij} is normalized if μ_i and μ_j are. The use of the max operator to solve severe disputes has no indication of involving the degree of certainty or the degree of fuzziness that each fuzzy assessment may carry.

Example: Assume that we have two pieces of evidence μ_i and μ_j. Let these pieces of evidence be committed to the same event $\theta \subset \Theta$. If $\mu_i(\theta) = 0.1$, $\mu_j(\theta) = 0.6$, and the threshold value $t = 0.4$, then Eq. 2.7 yields

$$\mu_{ij}(\theta) = \max(0.1, 0.6) = 0.6,$$

which means that the combined assessment is in favor of the greater one even if its degree of certainty is lower. This may not be acceptable if one considers that $\mu_i(\theta) = 0.1$ is more reliable than $\mu_j(\theta) = 0.6$. Another combination operator which is based on the

fuzzy set–theoretic symmetric sum was proposed by Silvert [73]:

$$\mu_{ij}(\theta) = \frac{g(\mu_i(\theta), \mu_j(\theta))}{g(\mu_i(\theta), \mu_j(\theta)) + g(1-\mu_i(\theta), 1-\mu_j(\theta))}, \tag{2.8}$$

where g is a continuous, non–decreasing, and non–negative function. If one assigns g to be the product, then

$$\mu_{ij}(\theta) = \frac{\mu_i(\theta)\mu_j(\theta)}{1-\mu_i(\theta)-\mu_j(\theta)+2\mu_i(\theta)\mu_j(\theta)}. \tag{2.9}$$

Using the same example with $\mu_i = 0.1$ and $\mu_j = 0.6$, we get $\mu_{ij} = 0.143$. This is an indication that this combination function judges in favor of the more certain assessment; *i.e.*, $\mu_i(\theta) = 0.1$ is considered more certain than $\mu_j(\theta) = 0.6$. However, if $\mu_i(\theta)$ is equal to 0, the combined assessment, $\mu_{ij}(\theta)$, is always equal to 0. This suggests that even if $\mu_j(\theta)$ is equal to 1, which is considered equally certain as $\mu_i(\theta)$, the combination is in favor of $\mu_i(\theta) = 0$. Along this argument, our desire is to use a combination criterion by which a test for certainty is performed.

Numerical Stability (Robustness): Fuzzy observations are not supposed to be exactly true; they are mathematically convenient realizations of often uncertain knowledge or belief. Numerical stability is an important issue in data fusion problems. A statistical inference is said to be numerically stable if a small error in an observation model should cause only a small error in the final conclusions [45]. The numerical stability of a system is often referred to as "robustness," which analyzes the insensitivity to small deviations from the assumption. Also, it is desirable that large deviations in an observation model should not cause complete failure. Robustness of a fusion function can also be defined in terms of smoothness as well. In other words, a fusion function is said to be robust if it is continuous and differentiable. Cheng and Kashyap [14] measured global robustness

by computing the upper bound on the absolute value of the slope of the combination formula with respect to its arguments. If the fusion function is differentiable, robustness is defined as

$$R = \sup \left| \frac{\partial f(\mu_i(\theta), \mu_j(\theta))}{\partial \mu_i(\theta)} \right| , \tag{2.10}$$

where sup is taken over the whole region of fusion. It can be shown that all combination rules with the associativity, commutativity, monotonicity, and continuity properties have unbounded R. Cheng and Kashyap stated [14]: "The quest for robust combination is often infeasible." This conclusion is based on the way they defined the robustness (*i.e.*, in terms of smoothness). However, one can choose a fusion function that is less sensitive to small changes than others.

Convexity: $\forall \theta \subset \Theta$, a combination of fuzzy memberships $\mu_1(\theta), \mu_2(\theta), \ldots, \mu_n(\theta)$ that is given by $\mu(\theta) = f(\mu_1, \mu_2, \ldots, \mu_n)$ is said to be convex if and only if

$$\min(\mu_1, \mu_2, \ldots, \mu_n) \leq \mu \leq \max(\mu_1, \mu_2, \ldots, \mu_n) . \tag{2.11}$$

The convexity of a combination is important because it allows for a compromise among pieces of evidence. Note that a subset of $\mu_1, \mu_2, \ldots, \mu_n$ is convex if, for all pairs of memberships in that subset, any convex combination

$$\mu(\theta) = f(\mu_i, \mu_j), \quad \forall i, j = 1, 2, \ldots, n$$

is also in the subset.

Idempotence: A combination function f is said to be idempotent if

$$\mu = \mu_1 = \mu_2 = \cdots = \mu_n,$$

then

$$f(\mu_1, \mu_2, \cdots, \mu_n) = \mu\,. \tag{2.12}$$

The idempotence property guarantees that if all information sources report the same support to an event, the outcome of data fusion must have the same support to that event.

Symmetry: A combination function f is said to be symmetric if

$$f(\mu_i, \mu_j) = f(\mu_j, \mu_i), \;\; \forall i, j = 1, 2, \ldots, n\,. \tag{2.13}$$

The symmetry property guarantees that the order of pieces of evidence in which they are received does not affect the final outcome of fusion.

Bisymmetry: A combination function f is said to be bisymmetric if the following combination of μ_1, μ_2, μ_3, and μ_4 holds:

$$f[f(\mu_1, \mu_2), f(\mu_3, \mu_4)] = f[f(\mu_1, \mu_3), f(\mu_2, \mu_4)]\,. \tag{2.14}$$

The bisymmetry property can be used as a substitute for associativity and also for symmetry (commutativity). Associativity is incompatible with the concept of weighted average [2, 22].

Monotonicity: The property of monotonicity guarantees the existence of a unique point (the identity) within the belief region, which divides the belief region into two segments: Values greater than the identity value are supportive to a given hypothesis. Values less than the identity value are non–supportive to a given hypothesis. Also, it guarantees that a stronger piece of evidence makes a stronger support [14]: $\forall i, j, k =$

$1, 2, \cdots, n$,

$$f(\mu_i, \mu_j) > f(\mu_i, \mu_k), \quad \forall \mu_j > \mu_k \,. \tag{2.15}$$

Independence: A set of information sources is said to be independent with respect to a hypothesis if the combined assessment on the hypothesis is a function of all fuzzy assessments committed to the same hypothesis generated by individual information sources in the set [14]. That is

$$\mu(\mu_1, \mu_2, \cdots, \mu_n) = f(\mu_1, \mu_2, \cdots, \mu_n)\,, \tag{2.16}$$

where n is the number of information sources.

Complementation [73]: This property assumes that the rule of combination is independent of whether we deal with sets or their complements:

$$1 - \mu(\mu_1, \mu_2, \cdots, \mu_n) = \mu(1 - \mu_1, 1 - \mu_2, \cdots, 1 - \mu_n)\,. \tag{2.17}$$

Regularity: A combination function is said to be a regular quazi–arithmetic mean if and only if the following conditions are satisfied [2]:

- $\mu(\mu_i, \mu_j) = l$ (idempotence).
- $\mu(\mu_i, \mu_j) = \mu(\mu_j, \mu_i)$ (symmetry).
- $\mu(\mu(\mu_1, \mu_2), \mu(\mu_3, \mu_4)) = \mu(\mu(\mu_1, \mu_3), \mu(\mu_2, \mu_4))$ (bisymmetry).
- $\mu(\mu_i, \mu_j) > \mu(\mu_i, \mu_k), \quad \forall \mu_j > \mu_k$ (monotonicity).

In the following, we seek the development of a new data fusion system in which the fusion formula satisfies all the desirable properties. All components of this system are derived in detail.

2.2 Data Fusion: A New Fusion Technique

Widely differing theories have been proposed for the combination of uncertain assessments. Theories are regarded differently either by the way they conceive uncertain data or by the adoption of a realist point of view. Most existing techniques for data fusion are unable to satisfy all the desirable properties stated earlier (see Appendix E). In this work, we regard the degree of certainty associated with each given assessment as the principal for the data fusion technique. We will also show that the new fusion function satisfies all the desirable properties, which are mentioned in the previous section. Let $\mu_1, \mu_2, \ldots, \mu_n$ be the truth values committed to a proposition $\theta \subset \Theta$, where $0 \leq \mu_i \leq 1, \quad \forall i = 1, 2, \ldots, n$. Because the truth region is defined within the interval $[0, 1]$, the identity of the truth region e is equal to 0.5. If $\mu_i = e$, then μ_i is considered maximally fuzzy and has zero degree of certainty. Then, μ_i is said to be more certain than μ_j if and only if $\forall i \neq j = 1, 2, \ldots, n$

$$|\mu_i - e| > |\mu_j - e| \ . \tag{2.18}$$

Equations 2.2 and 2.3 show that every assessment has a degree of certainty regardless of whether it is supportive or not.

The combined deviation of $\mu_1, \mu_2, \ldots, \mu_n$ from the maximally fuzzy point can be obtained by computing the average Minkowski Distance as follows:

$$d = \left[\frac{1}{n} \sum_{i=1}^{n} (\mu_i - e)^{\alpha} \right]^{1/\alpha} , \tag{2.19}$$

where α is odd; *i.e.*, $\alpha = 1, 3, 5, \ldots, \infty$, which represents a normalization factor that preserves the degree of certainty associated with each μ_i. From Eq. 2.19, the combined assessment can be established as the sum of the combined deviation from the identity,

d, and the maximally fuzzy point; *i.e.*,

$$\mu = e + \left[\frac{1}{n}\sum_{i=1}^{n}(\mu_i - e)^{\alpha}\right]^{1/\alpha}. \tag{2.20}$$

Since α is an odd, the value of $(\mu_i - e)^{\alpha}$ is negative, if μ_i is a non–supportive piece of evidence and is positive if μ_i is a supportive one. The parameter α determines the weight of each piece of evidence, which in turn expresses the relative importance of the aggregated bodies of evidence. This accommodation which is based on the degree of certainty is the core of this fusion technique. The desire is that the combination operator given by Eq. 2.20 can be defined to perform various strengths. That is, the Minkowski averaging operator can yield different aggregation functions, which are defined by different choices of the parameter $\alpha = 1, 3, \cdots, \infty$. Figure 2.2 shows a block diagram for the fusion formula component. In the following, we study the effect of different values of α on the fusion formula:

- For $\alpha = 1$, the combination function becomes

$$\mu = \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n},$$

  which coincides with the arithmetic mean.

- For $\alpha \geq 3$, (α is odd), a fuzzy assessment that has a higher degree of fuzziness will have less influence on the combined assessment.

- For $\alpha \rightarrow \infty$, the combination function performs the max operator; *i.e.*,

$$\lim_{\alpha \rightarrow \infty} \mu(\mu_1, \mu_2, \cdots, \mu_n) = \max(\mu_1, \mu_2, \cdots, \mu_n).$$

proof [51]: Let us assume, with no loss of generality, that $(\mu_i - e) = d_i$, and let $d_k > d_i, \ \forall i \neq k = 1, 2, \cdots, n$. Assume further that

$$q = \left[\frac{(\mu_1 - e)^\alpha + (\mu_2 - e)^\alpha + \cdots + (\mu_k - e)^\alpha + \cdots + (\mu_n - e)^\alpha}{n}\right]^{1/\alpha}.$$

Then,

$$\lim_{\alpha \to \infty} \ln q = \lim_{\alpha \to \infty} \frac{\ln(d_1^\alpha + d_2^\alpha + \cdots + d_k^\alpha + \cdots + d_n^\alpha)}{\alpha}.$$

Knowing that $d^\alpha = \exp(\alpha \ln d)$, and using l'Hospital's rule, we obtain:

$$\lim_{\alpha \to \infty} \ln q = \lim_{\alpha \to \infty} \frac{d_1^\alpha \ln d_1 + d_2^\alpha \ln d_2 + \cdots + d_k^\alpha \ln d_k + \cdots + d_n^\alpha \ln d_n}{d_1^\alpha + d_2^\alpha + \cdots + d_k^\alpha + \cdots + d_n^\alpha}$$

$$\lim_{\alpha \to \infty} \ln q = \lim_{\alpha \to \infty} \frac{(\frac{d_1}{d_k})^\alpha \ln d_1 + (\frac{d_2}{d_k})^\alpha \ln d_2 + \cdots + \ln d_k + \cdots + (\frac{d_n}{d_k})^\alpha \ln d_n}{(\frac{d_1}{d_k})^\alpha + (\frac{d_2}{d_k})^\alpha + \cdots + 1 + \cdots + (\frac{d_n}{d_k})^\alpha}$$

$$\lim_{\alpha \to \infty} \ln q = \ln d_k.$$

Or

$$\lim_{\alpha \to \infty} q = d_k.$$

Thus,

$$\lim_{\alpha \to \infty} \mu = e + \lim_{\alpha \to \infty} q = e + d_k.$$

Since $d_k = \mu_k - e$, hence

$$\lim_{\alpha \to \infty} \mu = \mu_k = \max(\mu_1, \mu_2, \cdots, \mu_n).$$

Therefore, for $\alpha = 1$, the operator is very soft and indicates perfect compromise among all degrees of memberships. On the other hand, for $\alpha \to \infty$, the operator is very hard, because it selects the largest degree of membership.

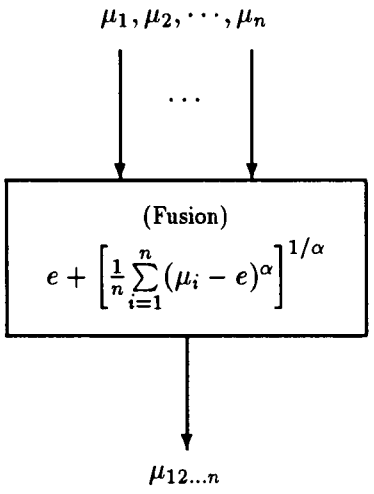


Figure 2.2: *A new data fusion function*

2.2.1 Test for the Desirable Properties

Generality: By inspection, Eq. 2.20 is able to combine any number of pieces of evidence simultaneously. It is also shown that Eq. 2.20 performs different operation strengths as α varies.

Test for Certainty: The degree of certainty of a measurement is based on how far the committed support is from the maximally fuzzy point. That is, the fuzzier θ is, the closer $\mu_i(\theta)$ is to the identity of the belief region, and the less fuzzy θ is, the closer $\mu_i(\theta)$ is to the nearest crisp value. Using the combination function given by Eq. 2.20, one can notice that a support that is closer to the identity value has a smaller effect on the outcome of the fusion. In contrast, a support that is closer to either crisp values has a greater effect on the outcome of the fusion. It is also noticeable that a piece of evidence may be supportive (*i.e.*, $\mu > e$) and in turn generates a positive difference when computing $(\mu - e)$ or non-supportive (*i.e.*, $\mu < e$) and in turn generates a nega-

tive difference when computing $(\mu - e)$. In this sense, the new fusion function tests for certainty and assigns a positive or negative additive support to each given assessment.

Numerical Robustness: Let L be the partial derivative of the combination function with respect to ϵ, where ϵ is a small change in a given assessment. To test for numerical robustness, we may compute L for different values of ϵ, then analyze the changes in the combined values. For $\mu_1, \mu_2, \cdots, \mu_n$, the combined assessment is given by

$$\mu(\mu_1, \mu_2, \cdots, \mu_n) = e + \left[\frac{(\mu_1 - e)^\alpha + (\mu_2 - e)^\alpha + \cdots + (\mu_n - e)^\alpha}{n}\right]^{1/\alpha}.$$

For $\mu_1 + \epsilon, \mu_2, \cdots, \mu_n$, the combined assessment is given by

$$\mu(\mu_1 + \epsilon, \mu_2, \cdots, \mu_n) = e + \left[\frac{(\mu_1 + \epsilon - e)^\alpha + (\mu_2 - e)^\alpha + \cdots + (\mu_n - e)^\alpha}{n}\right]^{1/\alpha}.$$

Because $L = \frac{\partial \mu}{\partial \epsilon}$, then

$$L = \frac{1}{n^{1/\alpha}}[(\mu_1 + \epsilon - e)^\alpha + (\mu_2 - e)^\alpha + \cdots + (\mu_n - e)^\alpha]^{(1-\alpha)/\alpha} \times (\mu_1 + \epsilon - e)^{\alpha-1}.$$

To test for numerical stability, we compute L for different values of ϵ. If a small error at the input level only causes a small change in L, then one can claim that the process is numerically stable.

Convex Combination: Because α is chosen to be an odd number, conflicting and/or non–conflicting assessments would show a compromise when they are fused, and therefore, the combination criteria given by Eq. 2.20 yields a convex combination.

proof: Let us assume that there are n assessments $\mu_1, \mu_2, \cdots, \mu_n$. Assume

$$\min(\mu_1, \mu_2, \cdots, \mu_n) = \mu_k$$

and

$$\max(\mu_1, \mu_2, \cdots, \mu_n) = \mu_l.$$

That is, $\forall i \neq k = 1, 2, \cdots, n$

$$(\mu_i - e)^\alpha \geq (\mu_k - e)^\alpha.$$

Also,

$$\sum_{i=1}^{n} \frac{(\mu_i - e)^\alpha}{n} \geq (\mu_k - e)^\alpha.$$

Or

$$\left[\sum_{i=1}^{n} \frac{(\mu_i - e)^\alpha}{n}\right]^{1/\alpha} \geq [(\mu_k - e)^\alpha]^{1/\alpha}.$$

Adding e to both sides of the equation and simplifying, we get:

$$e + \left[\sum_{i=1}^{n} \frac{(\mu_i - e)^\alpha}{n}\right]^{1/\alpha} \geq \mu_k.$$

Because the left hand side of the equation represents the combined assessment μ, which is given by Eq. 2.20, then

$$\mu \geq \mu_k$$

$$\mu \geq \min(\mu_1, \mu_2, \cdots, \mu_n).$$

Similarly, we can show that $\forall i \neq l = 1, 2, \cdots, n$,

$$\mu \leq \mu_l$$

$$\mu \leq \max(\mu_1, \mu_2, \cdots, \mu_n).$$

Therefore, $\forall\, \mu_i, i = 1, 2, \cdots, n$,

$$\min(\mu_1, \mu_2, \cdots, \mu_n) \leq \mu \leq \max(\mu_1, \mu_2, \cdots, \mu_n). \tag{2.21}$$

Then, $\mu(\mu_1, \mu_2, \cdots, \mu_n)$ expresses a compromise between conflicting and/or non-conflicting bodies of evidence, and therefore, it is a convex combination.

Idempotence Property: The combination function given by Eq. 2.20 is idempotent.

proof: Let $L = \mu_1 = \mu_2 = \cdots = \mu_n$. Applying Eq. 2.20 to compute the combined

assessment, we get

$$\mu = e + \left[\frac{(\mu_1 - e)^\alpha + (\mu_2 - e)^\alpha + \cdots + (\mu_n - e)^\alpha}{n}\right]^{1/\alpha} .$$

Or

$$\mu = e + \left[\frac{n(L - e)^\alpha}{n}\right]^{1/\alpha} = L . \tag{2.22}$$

Therefore, the combination function is idempotent.

Symmetry: The combination function given by Eq. 2.20 is symmetric.

proof: $\forall i, j = 1, 2, \cdots, n$, $\mu(\mu_i, \mu_j)$ and $\mu(\mu_j, \mu_i)$ can be computed using the combination formula given by Eq. 2.20, which yields

$$\mu(\mu_i, \mu_j) = e + \left[\frac{(\mu_i - e)^\alpha + (\mu_j - e)^\alpha}{2}\right]^{1/\alpha} , \tag{2.23}$$

and

$$\mu(\mu_j, \mu_i) = e + \left[\frac{(\mu_j - e)^\alpha + (\mu_i - e)^\alpha}{2}\right]^{1/\alpha} . \tag{2.24}$$

Because the right hand sides of both Eqs. 2.23 and 2.24 are equal, then

$$\mu(\mu_i, \mu_j) = \mu(\mu_j, \mu_i) . \tag{2.25}$$

The order in which μ_i and μ_j are received is irrelevant; hence the combination function is symmetric.

Bisymmetry: Assume that μ_1, μ_2, μ_3, and μ_4 are given assessments, and μ is the combination function given by Eq. 2.20. The combination function μ is bisymmetric because

$$\mu[\mu(\mu_1, \mu_2), \mu(\mu_3, \mu_4)] = \mu[\mu(\mu_1, \mu_3), \mu(\mu_2, \mu_4)] . \tag{2.26}$$

proof: Let the combination of μ_1 and μ_2 be $\mu(\mu_1, \mu_2) = d_1$, and the combination of μ_3 and μ_4 be $\mu(\mu_3, \mu_4) = d_2$. Applying the combination formula given by Eq. 2.20, we get

$$\mu(\mu_1, \mu_2) = e + \left[\frac{(\mu_1 - e)^\alpha + (\mu_2 - e)^\alpha}{2}\right]^{1/\alpha} = e + \left[\frac{m+n}{2}\right]^{1/\alpha} = d_1,$$

$$\mu(\mu_3, \mu_4) = e + \left[\frac{(\mu_3 - e)^\alpha + (\mu_4 - e)^\alpha}{2}\right]^{1/\alpha} = e + \left[\frac{k+l}{2}\right]^{1/\alpha} = d_2,$$

where m, n, k, and l are equal to $(\mu_1 - e)^\alpha$, $(\mu_2 - e)^\alpha$, $(\mu_3 - e)^\alpha$, and $(\mu_4 - e)^\alpha$, respectively. Then,

$$\mu[\mu(\mu_1, \mu_2), \mu(\mu_3, \mu_4)] = e + \left[\frac{(d_1 - e)^\alpha + (d_2 - e)^\alpha}{2}\right]^{1/\alpha}.$$

Substituting d_1 and d_2 into the last equation, we get

$$\mu[\mu(\mu_1, \mu_2), \mu(\mu_3, \mu_4)] = e + \left[\frac{\frac{m+n}{2} + \frac{k+l}{2}}{2}\right]^{1/\alpha}$$

$$= e + \left[\frac{m+n+k+l}{4}\right]^{1/\alpha}$$

$$= e + \left[\frac{(\mu_1 - e)^\alpha + (\mu_2 - e)^\alpha + (\mu_3 - e)^\alpha + (\mu_4 - e)^\alpha}{4}\right]^{1/\alpha}.$$

Then,

$$\mu[\mu(\mu_1, \mu_2), \mu(\mu_3, \mu_4)] = \mu(\mu_1, \mu_2, \mu_3, \mu_4). \tag{2.27}$$

Likewise, we can show that

$$\mu[\mu(\mu_1, \mu_3), \mu(\mu_2, \mu_4)] = \mu(\mu_1, \mu_2, \mu_3, \mu_4). \tag{2.28}$$

Equations 2.27 and 2.28 are the same. Thus, the combination function is bisymmetric. Note that the bisymmetry property has been used as a substitute for associativity and also for symmetry (commutativity) [2].

Monotonicity: $\forall i, j, k, = 1, 2, \cdots, n$, let μ_i, μ_j, and μ_k be the fuzzy assessments supporting an event $\theta \subset \Theta$. The combination function given by Eq. 2.20 is monotonic or

non–decreasing with respect to its arguments because

$$\mu(\mu_i, \mu_j) > \mu(\mu_i, \mu_k), \quad for \ \ \mu_j > \mu_k.$$

proof: Substituting the given assessments into Eq. 2.20, we get

$$\mu(\mu_i, \mu_j) = e + \left[\frac{(\mu_i - e)^\alpha + (\mu_j - e)^\alpha}{2}\right]^{1/\alpha},$$

and

$$\mu(\mu_i, \mu_k) = e + \left[\frac{(\mu_i - e)^\alpha + (\mu_k - e)^\alpha}{2}\right]^{1/\alpha},$$

where $\alpha = 1, 3, 5, \cdots, \infty$, and $\mu_j > \mu_k$. Because $\mu_j > \mu_k$, then

$$\frac{(\mu_j - e)^\alpha}{2} > \frac{(\mu_k - e)^\alpha}{2}.$$

Adding $(\mu_i - e)^\alpha/2$ to both sides of the equation, we get:

$$\frac{(\mu_i - e)^\alpha}{2} + \frac{(\mu_j - e)^\alpha}{2} > \frac{(\mu_i - e)^\alpha}{2} + \frac{(\mu_k - e)^\alpha}{2}.$$

Then,

$$\left[\frac{(\mu_i - e)^\alpha}{2} + \frac{(\mu_j - e)^\alpha}{2}\right]^{1/\alpha} > \left[\frac{(\mu_i - e)^\alpha}{2} + \frac{(\mu_k - e)^\alpha}{2}\right]^{1/\alpha}.$$

Adding e to both sides yields

$$\mu(\mu_i, \mu_j) > \mu(\mu_i, \mu_k), \quad \forall \mu_j > \mu_k \,. \tag{2.29}$$

Complementation: Using the combination function given by Eq. 2.20 shows that dealing with given assessments is equivalent to dealing with their complement:

$$1 - \mu(\mu_1, \mu_2, \cdots, \mu_n) = \mu(1 - \mu_1, 1 - \mu_2, \cdots, 1 - \mu_n).$$

proof: From Eq. 2.20, we have

$$\mu(\mu_1, \mu_2, \cdots, \mu_n) = e + \left[\frac{(\mu_1 - e)^\alpha + (\mu_2 - e)^\alpha + \cdots + (\mu_n - e)^\alpha}{n}\right]^{1/\alpha}.$$

Then,

$$\mu(1-\mu_1, 1-\mu_2, \cdots, 1-\mu_n) = e+ \left[\frac{(1-\mu_1-e)^\alpha + (1-\mu_2-e)^\alpha + \cdots + (1-\mu_n-e)^\alpha}{n}\right]^{1/\alpha}$$

$$\mu(1-\mu_1, 1-\mu_2, \cdots, 1-\mu_n) = e- \left[\frac{(\mu_1-e)^\alpha + (\mu_2-e)^\alpha + \cdots + (\mu_n-e)^\alpha}{n}\right]^{1/\alpha} . \quad (2.30)$$

Using Eq. 2.20, we can write

$$1-\mu(\mu_1, \mu_2, \cdots, \mu_n) = e- \left[\frac{(\mu_1-e)^\alpha + (\mu_2-e)^\alpha + \cdots + (\mu_n-e)^\alpha}{n}\right]^{1/\alpha} . \quad (2.31)$$

Comparing Eqs. 2.30 and 2.31, we see that

$$\mu(1-\mu_1, 1-\mu_2, \cdots, 1-\mu_n) = 1-\mu(\mu_1, \mu_2, \cdots, \mu_n) . \quad (2.32)$$

Similarly, if the two sources of information are in extreme dispute, *i.e.*, $\mu_i = x$ and $\mu_j = 1-x$, the combination is given by

$$\mu(x, 1-x) = e.$$

Regularity: Since the following conditions

- $\mu(l,l) = l$ (idempotence),
- $\mu(\mu_i, \mu_j) = \mu(\mu_j, \mu_i)$ (symmetry),
- $\mu[\mu(\mu_1,\mu_2), \mu(\mu_3,\mu_4)] = \mu[\mu(\mu_1,\mu_3), \mu(\mu_2,\mu_4)]$ (bisymmetry), and
- $\mu(\mu_i, \mu_j) > \mu(\mu_i, \mu_k), \quad \forall \mu_j > \mu_k$ (monotonic)

are satisfied, then the combination function given by Eq. 2.20 is a regular quazi–arithmetic mean [2].

2.3 The New Fusion Technique vs. Other Methods

We tested the new fusion formula against several desirable properties. We showed that the new fusion formula relies on the degree of certainty or the degree of fuzziness when computing the combined assessment. In other words, it tests for the reliability of the information to be fused and it combines all the given assessments through a reconciliation process. In light of the desirable properties, we investigate several methods that already exist in the literature.

2.3.1 The Super Bayesian Approach

A complete derivation of the Super Bayesian Approach (SBA) is given in Appendix A. In the following, we test the SBA against the desirable properties.

Generality: The SBA is a direct application of Bayes' theorem [7]:

$$p(\theta_i|x) = \frac{p(x|\theta_i) \cdot p(\theta_i)}{\sum_{j=1}^{n} p(x|\theta_j) \cdot p(\theta_j)},$$

where $i = 1, 2, \ldots, n$, so that for every subscript of θ, we would have a related posterior probability associated with it. Hence, the SBA is able to combine any number of pieces of evidence simultaneously.

Test for Certainty: Bayesians assume the availability of detailed prior opinions. In this framework, a proposition is a subset of a set over which priors are defined. This would give priors greater influence on the outcome of the aggregation even though priors are sometimes assumed with high degrees of doubt. The SBA does not fairly treat the old evidence and the new. It considers the new evidence as a proposition by which the old evidence (prior probability) is conditioned. When applying the SBA, it becomes a

special case of Bayes' theory, and three cases of aggregation are observed: (1) If the two sources of information are in conflict, *i.e.*, one of them is supportive ($p_1 > 0.5$) and the other is not supportive ($p_2 < 0.5$), the method performs a compromise between p_1 and p_2. (2) If both sources of information are supportive, the method performs an overestimated combined assessment. (3) If both sources of information are not supportive, the method performs an underestimated combined assessment. In this sense, there is a test for certainty between given pieces of evidence using the SBA. This analysis is true only with the interval $0 < p_i < 1$, $\forall i$.

Numerical Robustness: Let $p_1(\theta)$ and $p_2(\theta)$ be two probabilities reported by two information sources in support of the occurrence of an event $\theta \subset \Theta$. The combined probability can be computed using the SBA as follows:

$$Comb(p_1, p_2) = \frac{p_1 p_2}{p_1 p_2 + (1 - p_1)(1 - p_2)}$$

Let p_1 be incremented by a small value ϵ. The combined probability becomes

$$Comb(p_1 + \epsilon, p_2) = \frac{(p_1 + \epsilon)p_2}{(p_1 + \epsilon)p_2 + (1 - p_1 - \epsilon)(1 - p_2)}$$

The derivative of the combined probability with respect to ϵ is

$$L = \frac{\partial Comb}{\partial \epsilon} = \frac{p_2(1 - p_2)}{[1 + (2p_2 - 1)(p_1 + \epsilon) - p_2]^2}.$$

By inspection, small changes in the values of ϵ will not cause drastic changes in L. This analysis is restricted within the interval $0 < p_i < 1$, $\forall i$. Hence, the SBA is robust.

Convexity: From Appendix D, we have shown that if p_1 and p_2 are greater than 0.5, they are in relative agreement, then

$$Comb(p_1, p_2) > \max(p_1, p_2).$$

If p_1 and p_2 are less than 0.5, they are in relative agreement; then

$$Comb(p_1, p_2) < \min(p_1, p_2).$$

If one of the probabilities is greater than 0.5 and the other is less than 0.5, they are in disagreement; then

$$\min(p_1, p_2) < Comb(p_1, p_2) < \max(p_1, p_2).$$

Therefore, the SBA performs a convex combination only if the sources of information are in conflict.

Idempotence: The SBA is not idempotent. That is,

$$Comb(x, x) \neq x.$$

proof: Let us assume that $p_1(\theta) = p_2(\theta) = x$. Applying the SBA, the combined probability is

$$\begin{aligned} Comb(p_1, p_2) &= \frac{p_1 p_2}{p_1 p_2 + (1-p_1)(1-p_2)} \\ &= \frac{x^2}{x^2 + (1-x)(1-x)} = \frac{x^2}{2x^2 - 2x + 1} \\ &\neq x\,. \end{aligned}$$

Then, the SBA is not idempotent except at the crisp values or at the midpoint.

Symmetry: Using the SBA, the combined support is

$$Comb(p_1, p_2) = \frac{p_1 p_2}{p_1 p_2 + (1-p_1)(1-p_2)}. \tag{2.33}$$

This expression is symmetric; hence, the SBA is symmetric.

Bisymmetry: The SBA is not bisymmetric because

$$Comb[Comb(p_1, p_2), Comb(p_3, p_4)] \neq Comb[Comb(p_1, p_3), Comb(p_2, p_4)]$$

proof: Let p_1, p_2, p_3, and p_4 be four probability numbers reported by four sources of information in support of an event $\theta \subset \Theta$. Using the SBA, we can write

$$Comb(p_1, p_2) = \frac{p_1 p_2}{p_1 p_2 + (1 - p_1)(1 - p_2)},$$

$$Comb(p_3, p_4) = \frac{p_3 p_4}{p_3 p_4 + (1 - p_3)(1 - p_4)},$$

$$Comb(p_1, p_3) = \frac{p_1 p_3}{p_1 p_3 + (1 - p_1)(1 - p_3)},$$

and

$$Comb(p_2, p_4) = \frac{p_2 p_4}{p_2 p_4 + (1 - p_2)(1 - p_4)}.$$

Then,

$$Comb[Comb(p_1, p_2), Comb(p_3, p_4)] =$$

$$\frac{p_1 p_2 p_3 p_4}{2p_1 p_2 p_3 p_4 + p_1 p_2 + p_3 p_4 - p_1 p_2 p_3 - p_1 p_2 p_4 - p_1 p_3 p_4 - p_2 p_3 p_4}, \tag{2.34}$$

$$Comb[Comb(p_1, p_3), Comb(p_2, p_4)] =$$

$$\frac{p_1 p_2 p_3 p_4}{2p_1 p_2 p_3 p_4 + p_2 p_4 + p_1 p_3 - p_1 p_2 p_3 - p_1 p_2 p_4 - p_1 p_3 p_4 - p_2 p_3 p_4}, \tag{2.35}$$

By comparison, the numerators of the right hand sides of Eqs. 2.34 and 2.35 are equal, but the denominators are not. Because the right hand sides of Eqs. 2.34 and 2.35 are not equal, the SBA is not bisymmetric.

Monotonicity: Let $p_1(\theta)$ and $p_2(\theta)$ be two probabilities reported by two information sources in support of an event $\theta \subset \Theta$. Using the SBA, the combination of p_1 and p_2 is given by

$$f(p_1, p_2) = \frac{p_1 p_2}{p_1 p_2 + (1 - p_1)(1 - p_2)}.$$

Because $f_{p_1} \geq 0 \ \forall p_2$ and $f_{p_2} \geq 0$, f is monotonic $\forall p_1$.

proof:

$$f_{p_1} = \frac{p_2 - p_2^2}{(1 + 2p_1p_2 - p_1 - p_2)^2}.$$

Because $0 \leq p_2 \leq 1$ and $(1 + 2p_1p_2 - p_1 - p_2)^2 \geq 0$, hence $f_{p_1} \geq 0$, $\forall p_2$. Because f is symmetric, $f_{p_2} \geq 0$, $\forall p_1$.

Complementation: The SBA satisfies the property of complementation. That is,

$$1 - Comb(p_1, p_2) = Comb(1 - p_1, 1 - p_2)$$

proof:

$$1 - Comb(p_1, p_2) = \frac{1 - p_1 - p_2 + p_1p_2}{p_1p_2 + (1 - p_1)(1 - p_2)},$$

and

$$Comb(1 - p_1, 1 - p_2) = \frac{1 - p_1 - p_2 + p_1p_2}{p_1p_2 + (1 - p_1)(1 - p_2)}.$$

Consequently, the SBA satisfies the property of complementation.

Regularity and the SBA: A combination function is said to be a regular–quazi arithmetic mean if it is idempotent, symmetric, bisymmetric, and monotonic [2]. Because the SBA is not idempotent, it is not a regular–quazi arithmetic mean.

2.3.2 Dempster's Rule of Combination

In the following, we test Dempster's rule of combination against the desirable properties.

Limitations in Computation Efficiency: Let $\mu_1, \mu_2, \cdots, \mu_n$ be a collection of belief functions committed to a hypothesis $\theta \subset \Theta$. To combine several belief functions, one must repeatedly apply Dempster's rule of combination to perform pairwise orthogonal

sums. For $n = 4$, the resulting orthogonal sum (the combined belief) is computed as follows:

$$\mu_1 \oplus \mu_2$$

$$(\mu_1 \oplus \mu_2) \oplus \mu_3$$

$$((\mu_1 \oplus \mu_2) \oplus \mu_3) \oplus \mu_4.$$

In practice, a pairwise procedure may cause lengthy computations, in particular when the number of given belief functions is large. From the previous example, one can deduce that in combining n belief functions, $n - 1$ orthogonal sum computations are required.

Convexity: Dempster's rule of combination yields a non–convex combination and shows no compromise between degrees of belief. More precisely, the combined assessment is always overestimated; *i.e.*,

$$Comb(\mu_1, \mu_2) \geq \max(\mu_1, \mu_2)\,. \tag{2.36}$$

proof: Let μ_1 and μ_2 be two belief functions committed to an event $\theta \subset \Theta$. Using Dempster's rule of combination, the total support committed to θ is given by

$$Comb(\mu_1, \mu_2) = 1 - (1 - \mu_1)(1 - \mu_2) = \mu_1 + \mu_2 - \mu_1\mu_2\,. \tag{2.37}$$

Because μ_1 and μ_2 are within the unit interval $[0, 1]$,

$$\mu_1\mu_2 \leq \min(\mu_1, \mu_2)$$

or

$$-\mu_1\mu_2 \geq -\min(\mu_1, \mu_2)\,. \tag{2.38}$$

Adding $\mu_1 + \mu_2$ to both sides of Eq. 2.38, we get

$$\mu_1 + \mu_2 - \mu_1\mu_2 \geq \mu_1 + \mu_2 - \min(\mu_1, \mu_2)\,. \tag{2.39}$$

From Eqs. 2.37 and 2.39, we have

$$Comb(\mu_1, \mu_2) \geq \mu_1 + \mu_2 - \min(\mu_1, \mu_2)\,. \tag{2.40}$$

Since $\mu_1 + \mu_2 = \max(\mu_1, \mu_2) + \min(\mu_1, \mu_2)$, hence

$$Comb(\mu_1, \mu_2) = \max(\mu_1, \mu_2) + \min(\mu_1, \mu_2) - \min(\mu_1, \mu_2) = \max(\mu_1, \mu_2).$$

Therefore, for any values of μ_1 and μ_2, Eq. 2.36 is true.

Test for Certainty: Dempster's rule performs a non–convex combination, and it overestimates the combined evidence. Information sources that are non–supportive are basically ignored if at least one supportive information source exists. However, if there is a large number of non–supportive information sources, the combined assessment is a supportive one regardless of whether or not a supportive information source exists. Therefore, there is no test on the degree of certainty that information sources may carry. Instead, non–supportive belief functions are suppressed, and they are given no influence over the combined assessment.

Robustness [22]: Let a, b, and c be three mutually exclusive and exhaustive hypotheses. Assume that each hypothesis is supported by two information sources as follows:

$$\begin{aligned}
m_1(a) &= \mu_1(a) - \epsilon, \quad m_2(a) = \epsilon \\
m_1(b) &= \mu_1(b) - \epsilon, \quad m_2(b) = \mu_2(b) \\
m_1(c) &= \epsilon, \qquad\quad\; m_2(c) = \mu_2(c) - \epsilon\,.
\end{aligned}$$

For this illustration, let

$$\mu_1(a) = \mu_2(c) = 0.99$$

and

$$\mu_1(b) = \mu_2(b) = 0.01.$$

Using Dempster's rule of combination, one can compute the total belief committed to b as follows:

$$m(b) = \frac{0.0001}{0.0001 + 2\epsilon(0.99 - \epsilon)}$$

To test for numerical stability, we compute the derivative of $m(b)$ with respect to the small deviation ϵ [43]; *i.e.*,

$$\frac{dm(b)}{d\epsilon} = \frac{(0.0001)\ 2\ (0.99 - 2\epsilon)}{[0.0001 + 2\epsilon(0.99 - \epsilon)]^2}.$$

To test the numerical sensitivity of the new fusion formula against Dempster's rule of combination, one can construct a table that contains the values of

$$\frac{dm(b)}{d\epsilon}$$

for different values of ϵ (see Table 2.2). As a result, Dempster's rule of combination is more sensitive to small changes than the fusion function. Dempster's rule of combination was shown to be a non–robust aggregation tool for combining evidence [22, 43, 84].

Idempotence: Dempster's rule of combination is not idempotent. That is,

$$m(x, x) \neq x.$$

proof: Let us assume that $\mu_1(\theta) = \mu_2(\theta) = x$. Applying Dempster's rule of combination, the combined support is

$$m_{12}(\theta) \quad = \quad 1 - (1 - x)(1 - x)$$

Table 2.2: *Values of the derivative for different values of* ϵ

ϵ	$\frac{dm(b)}{d\epsilon}$	$\frac{d\mu(b)}{d\epsilon}$
0.1	0.005	0.06
0.01	0.5	0.02
0.001	45.76	0.016631
0.0001	2230	0.01663
0.00001	13796	0.0166
0.0	19800	0.0165

$$= \quad 1 - (1 - 2x + x^2) = 2x - x^2 = x(2 - x)$$

$$\neq \quad x\,.$$

Dempster's rule of combination is not idempotent except at the crisp values, where $x = 0$ or $x = 1$.

Symmetry: Dempster's rule of combination is symmetric. That is, the order in which pieces of evidence are received is irrelevant.

proof: Let μ_1 and μ_2 be two belief functions committed to an event $\theta \subset \Theta$. Using Dempster's rule of combination, the total support committed to θ is given by

$$Comb(\mu_1, \mu_2) = 1 - (1 - \mu_1)(1 - \mu_2) = \mu_1 + \mu_2 - \mu_1\mu_2 = Comb(\mu_2, \mu_1)\,. \qquad (2.41)$$

Bisymmetry: Dempster's rule of combination is not bisymmetric.

proof: Let μ_1, μ_2, μ_3, and μ_4 be four belief functions committed to an event $\theta \subset \Theta$. Using Dempster's rule of combination, we can write

$$Comb(\mu_1, \mu_2) = 1 - (1 - \mu_1)(1 - \mu_2) = \mu_1 + \mu_2 - \mu_1\mu_2\,. \qquad (2.42)$$

$$Comb(\mu_3, \mu_4) = 1 - (1 - \mu_3)(1 - \mu_4) = \mu_3 + \mu_4 - \mu_3\mu_4 \,. \tag{2.43}$$

Then,

$$Comb[Comb(\mu_1, \mu_2), Comb(\mu_3, \mu_4)] = \mu_1 + \mu_2 - \mu_1\mu_2 + \mu_3 + \mu_4$$

$$-\mu_3\mu_4 + \mu_1\mu_3 + \mu_1\mu_4 - \mu_1\mu_3\mu_4 + \mu_2\mu_3 + \mu_2\mu_4$$

$$-\mu_2\mu_3\mu_4 - \mu_1\mu_2\mu_3 - \mu_1\mu_2\mu_4 + \mu_1\mu_2\mu_3\mu_4 \,. \tag{2.44}$$

Also,

$$Comb(\mu_1, \mu_3) = 1 - (1 - \mu_1)(1 - \mu_3) = \mu_1 + \mu_3 - \mu_1\mu_3 \,. \tag{2.45}$$

$$Comb(\mu_2, \mu_4) = 1 - (1 - \mu_2)(1 - \mu_4) = \mu_2 + \mu_4 - \mu_2\mu_4 \,. \tag{2.46}$$

Then,

$$Comb[Comb(\mu_1, \mu_3), Comb(\mu_2, \mu_4)] = \mu_1 + \mu_3 - \mu_1\mu_3 + \mu_2 + \mu_4$$

$$-\mu_2\mu_4 + \mu_1\mu_2 + \mu_1\mu_4 - \mu_1\mu_2\mu_4 + \mu_2\mu_3 + \mu_3\mu_4$$

$$-\mu_2\mu_3\mu_4 - \mu_1\mu_2\mu_3 - \mu_1\mu_3\mu_4 + \mu_1\mu_2\mu_3\mu_4 \,. \tag{2.47}$$

Computing the difference, d, between the right hand sides of Eqs. 2.44 and 2.47, we get

$$d = 2(\mu_1 - \mu_4)(\mu_3 - \mu_2).$$

The latter equation indicates that d is not equal to 0 except for the trivial cases where $\mu_3 = \mu_2$ or $\mu_1 = \mu_4$. Hence,

$$Comb[Comb(\mu_1, \mu_2), Comb(\mu_3, \mu_4)] \neq Comb[Comb(\mu_1, \mu_3), Comb(\mu_2, \mu_4)].$$

Therefore, Dempster's rule of combination is not bisymmetric.

Monotonicity: Let μ_1, μ_2, and μ_3 be three belief functions committed to an event $\theta \subset \Theta$, where $\mu_3 > \mu_2$. Then,

$$Comb(\mu_1, \mu_2) = 1 - (1 - \mu_1)(1 - \mu_2) = \mu_1 + \mu_2 - \mu_1\mu_2\,, \tag{2.48}$$

and

$$Comb(\mu_1, \mu_3) = 1 - (1 - \mu_1)(1 - \mu_3) = \mu_1 + \mu_3 - \mu_1\mu_3\,. \tag{2.49}$$

Subtracting Eq. 2.49 from Eq. 2.48, we get

$$Comb(\mu_1, \mu_2) - Comb(\mu_1, \mu_3) = (1 - \mu_1)(\mu_2 - \mu_3).$$

Because $\mu_1, \mu_2 \in [0, 1]$ and $\mu_3 > \mu_2$,

$$(1 - \mu_1) \geq 0,$$

and

$$(\mu_2 - \mu_3) < 0.$$

Therefore,

$$Comb(\mu_1, \mu_2) - Comb(\mu_1, \mu_3) \leq 0 \quad for \quad \mu_3 > \mu_2.$$

Or

$$Comb(\mu_1, \mu_3) \geq Comb(\mu_1, \mu_2) \quad for \quad \mu_3 > \mu_2.$$

Dempster's rule of combination is monotonic.

Complementation: Let μ_1 and μ_2 be two belief functions committed to an event $\theta \subset \Theta$,

$$1 - Comb(\mu_1, \mu_2) = 1 + \mu_1\mu_2 - \mu_1 - \mu_2\,, \tag{2.50}$$

Table 2.3: *Summary of combination techniques against the desirable properties*

Properties	Dempster's rule	The Super Bayesian	The Fusion Formula
Generality	No	Yes	Yes
Certainty Test	No	Yes	Yes
Convexity	No	–	Yes
Robustness	No	Yes	Yes
Idempotence	No	No	Yes
Symmetry	Yes	Yes	Yes
Bisymmetry	No	No	Yes
Monotonicity	Yes	Yes	Yes
Complementation	No	Yes	Yes
Regularity	No	No	Yes

$$Comb(1-\mu_1, 1-\mu_2) = 1 - \mu_1\mu_2\,. \tag{2.51}$$

Computing the difference of Eqs. 2.50 and 2.51, we get

$$Comb(1-\mu_1, 1-\mu_2) - [1 - Comb(\mu_1, \mu_2)] = \mu_1 + \mu_2 - 2\mu_1\mu_2 \neq 0.$$

Hence, dealing with pieces of evidence is not equivalent to dealing with their complements, except at the crisp values where $\mu_1 = \mu_2 = 1$, or $\mu_1 = \mu_2 = 0$.

Regularity: Because Dempster's rule of combination is not idempotent, hence it is not a regular–quazi arithmetic mean [2].

A summary of testing the new fusion formula, the SBA, and Dempster's rule of combination against all desirable properties is given in Table 2.3.

2.4 Fuzzification: Modeling Uncertainty

Fuzzification or modeling uncertainty is the first component in our data fusion system. Before fusion, all types of data are fuzzified. Generally, the process of fuzzification is a mechanism by which one can transform a nonfuzzy set to a fuzzy set or increase the fuzziness of a fuzzy set [88]. Nonfuzzy sets can be fuzzified by assigning to each subset in the universe of discourse a fuzzy value. The fuzzification process must be carried out before any attempt to fuse data because (1) various information sources often report different types of data; after fuzzification, a common domain for all types of data can be established. (2) To apply fuzzy mathematics, it is appropriate to transform nonfuzzy sets to fuzzy sets. (3) It is also a way of simplifying tasks where a high degree of accuracy is neither needed nor required [87]. Under the assumption of uncertainty associated with sources of information, measurements may be fuzzified by assigning to them fuzzy values such as strong, weak, true, untrue, very true, etc. Even though these linguistic restrictions do not have sharp boundaries, they allow for numerical computations. Hence, our goal is to fuzzify different types of data in a fuzzy framework such that we will be able to employ fuzzy logic or reasoning. Such procedures include operators by which decisions can be deduced from fuzzy assessments. This framework is viewed as more realistic and closer to human reasoning than the traditional two–valued logic [84]. In the context of fuzzy reasoning, members of a fuzzy subset are assigned degrees of compatibility or membership. The degree of membership characterizes the confidence measure assigned to a given assertion. Each fuzzy subset in the universe of discourse is assigned degrees of membership with respect to a fuzzy concept in the same universe. In the following, we will introduce a method for fuzzy modeling through the construc-

tion of membership functions. Also, a complete fuzzification scheme for universes of discourse will be introduced.

2.4.1 Membership Functions

The concept of membership function was introduced by Zadeh [84, 85]. It provides a means to characterize vagueness where the proposition, "belongingness," of a subset to a set can take a degree of membership within the unit interval $[0, 1]$, where the degree $\mu(\theta) = 1$ represents a full belongingness, degrees $0 < \mu(\theta) < 1$ represent intermediate belongingness, and the degree $\mu(\theta) = 0$ represents a full non–belongingness. Formally,

$$\forall\theta \subset \Theta, \ \ \mu_A : \Theta \rightarrow [0, 1],$$

where μ_A is the membership function which maps from Θ to the membership space $[0, 1]$ and $\mu_A(\theta)$ is the grade of membership of θ in A. For a given fuzzy term, membership functions can have different shapes depending on the designer's preference or experience [52]. The construction of membership functions is often based on perceptions of fuzzy labels. Experts use membership functions to reflect their opinions regarding a fuzzy concept. Due to the nature of fuzziness, we would expect that experts do not perceive a subjective attribute identically [41]. It follows that the construction of membership functions has no specific rules. In the early development of the fuzzy sets theory, membership functions were constructed using standard functions such as a bell–shaped function and/or monotonically increasing and decreasing S–shaped functions. These functions are convenient to fit approximately specific membership functions [89]. In practice, however, many fuzzy logic researchers use triangular and trapezoidal shapes in order to simplify computations [52]. In the following, triangular and trapezoidal shapes are used to represent membership functions.

2.4.2 Representation of Fuzzy Values

Let an input be represented by a numerically valued variable which ranges over the interval $[0, 1]$. This interval is regarded as the universe of discourse Θ. Linguistic variables such as *weak, strong*, etc., are interpreted as values for fuzzy restrictions on some partitioned values of the universe of discourse. These variables are fuzzy concepts which are fuzzy subsets of Θ. The fuzzy subsets are characterized by compatibility functions such as μ_{weak}, μ_{strong}, etc. The compatibility functions are the degrees to which numerical intensities, say $\theta = 0.3$ and $\theta = 1.0$, are compatible with a fuzzy concept. For instance,

$$\mu_{weak}(0.3) = 0.7,$$

and

$$\mu_{weak}(1.0) = 0.2.$$

In the context of approximate reasoning, μ characterizes the confidence measure of assertion. That is, $\mu_{weak}(0.3) = 0.7$ means that the degree of confidence of $\theta = 0.3$ as a weak input is 0.7. Thus, a fuzzy subset is characterized by a membership function, μ, over Θ:

$$\Theta = (\theta_1, \theta_2, \ldots, \theta_n)$$

$$\mu : \Theta \rightarrow [0, 1].$$

The fuzzy concept,"weak," can be represented by a membership function that characterizes a weak portion of the base variable (*i.e.*, the universe of discourse). Hence, suppose that we quantize the universe of discourse as *weak, moderate*, and *strong*. This will lead to three fuzzy–set values:

Weak: **WE**

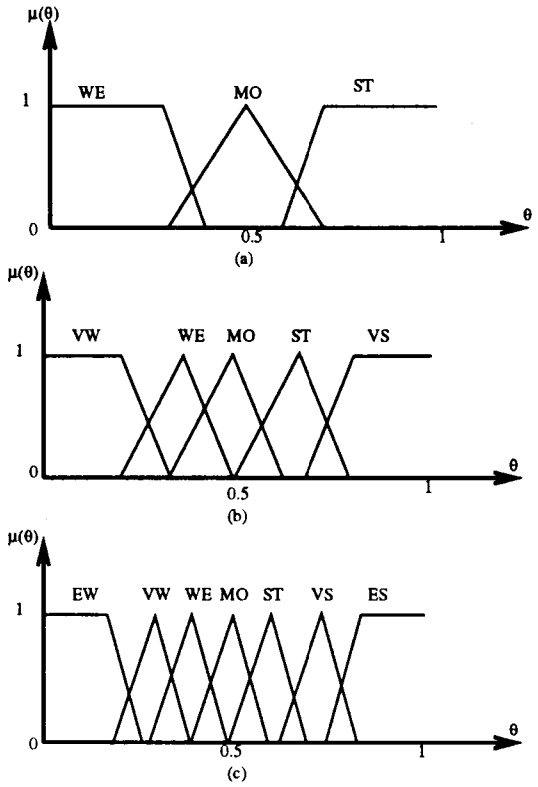


Figure 2.3: *A modeling example: (a) three, (b) five, and (c) seven fuzzy values.*

Moderate: **MO**

Strong: **ST**

If three fuzzy values are unable to adequately characterize the universe of discourse, then more fuzzy values are needed. Figure 2.3 shows membership function graphs for the fuzzy subsets of a universes of discourse, which is quantized into three, five, and seven fuzzy values. We use the fuzzy values *very weak, very strong, extremely weak,* and *extremely strong*, which are abbreviated by **VW, VS, EW,** and **ES**, respectively. Sensitivity of data and application domain play a major role in the selection of fuzzy values and the quantization procedure as well. The fuzzification process is best illustrated by an example. Assume that the universe of discourse represents an edge map extracted

from a video image. Suppose that the universe of discourse is quantized into the three fuzzy values **WE, MO,** and **ST**. For each input value in the universe of discourse, we assign a fit vector of length 3, where each element in the fit vector is assigned by the corresponding membership function. That is, for an input value $\theta \subset \Theta$, we have

$$\theta \longrightarrow [\mu_{we}(\theta), \mu_{mo}(\theta), \mu_{st}(\theta)].$$

These values can be determined by convolving a Kronecker delta function centered at the point value θ with each of the three fuzzy subsets [52]. Figure 2.4 shows the output of this fuzzification. The original video edge map is quantized into three clusters, each of which is characterized by a fuzzy value. The intensity of each cluster indicates the cluster's compatibility to the assigned fuzzy value. If we have three sources of information, say S_1, S_2, and S_3, gathering information from the same scene, each source would independently generate its own assertion. At a location x in every input feature map, we define a universe of discourse with the following assertion:

$$S_i \text{ reports that } x \text{ is } \theta_i$$

The associated membership functions with "x is θ_i", where $\theta_i \in \Theta_i$ are

$$\mu_{we}(x \ is \ \theta_i), \quad \mu_{mo}(x \ is \ \theta_i), \ and \quad \mu_{st}(x \ is \ \theta_i),$$

where $i = 1, 2, \cdots, n$ is the number of information sources. Therefore, for every sensory measurement, there are three overlapped fuzzy values. These fuzzy values allow us to encode structured knowledge in a numerical framework. In this sense, this fuzzification procedure is able to represent uncertain sensory measurements.

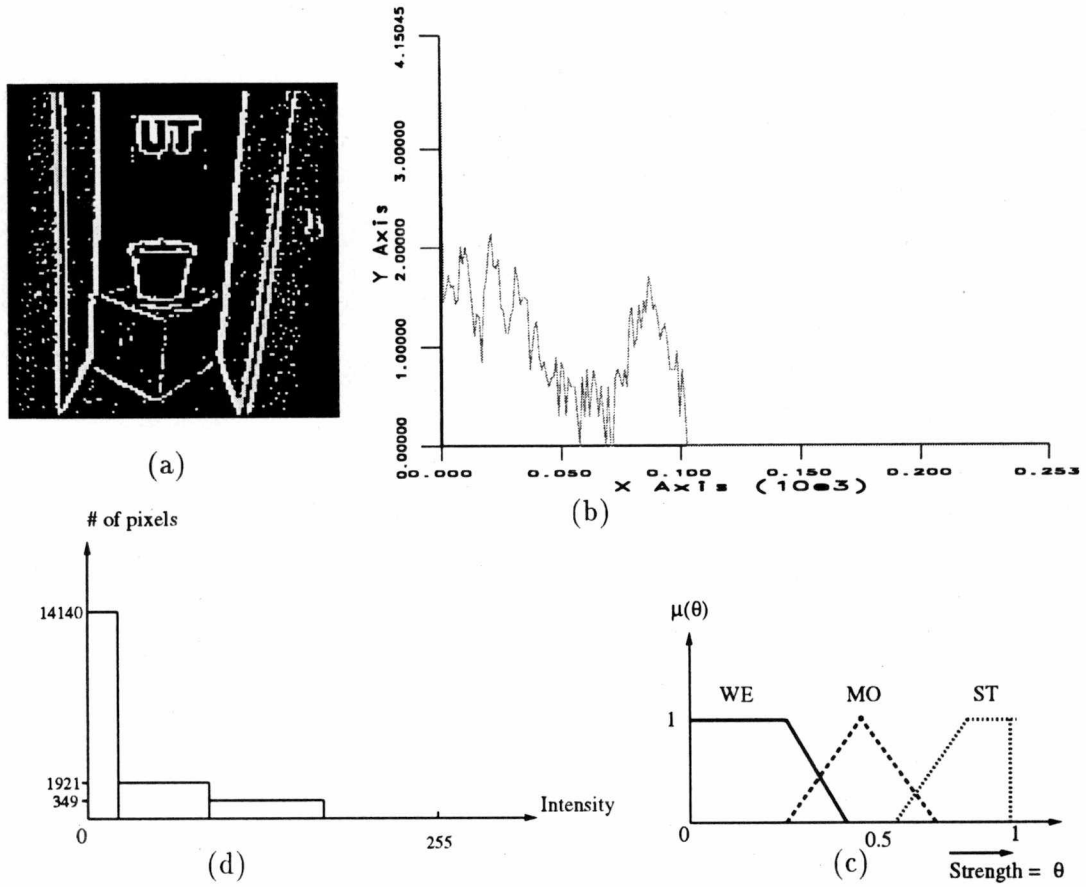


Figure 2.4: *Fuzzification: (a) video edge map; (b) histogram for (a) plotted as the ordinate to a logarithmic scale; (c) an approximated histogram for (a); (c) fuzzy-set values.*

2.5 Fuzzy Reasoning: Data Fusion through Fuzzy Rules Combination

Propositional rules with logical implications enhance the capability of a system to infer information. In this system, structured knowledge is encoded (*i.e.*, fuzzified) in a numerical framework. Each fuzzy label defines a numerical portion of the universe of discourse. Fuzzy reasoning is governed by rules that are sufficient to generate consensus among fuzzy entries. We construct a rule base (RB) that is able to reflect subjective reasoning in particular when conditional statements are involved. Elements of the RB

are found according to the application domain and the nature of the problem. The RB asserted through a mapping scheme that is able to map input values with different domains to an output value with only one domain. A mapping function $F : \theta_1 \oplus \theta_2 \oplus \theta_3 \longrightarrow \theta$ maps each element of the domain $\theta_1 \oplus \theta_2 \oplus \theta_3$ to only one element of θ. This mapping function performs a logical implication from which we construct the RB. Fuzzy logic engineers, therefore, define F to articulate their opinions regarding the outcome of an inference process. For instance,

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(a_1, a_2, a_3) \longrightarrow a\} = \text{elements of } \theta,$$

where a_1, a_2, a_3, and a are prespecified linguistic values within the universe(s) of discourse. In our case, they are the quantized segments *weak*, *strong*, etc. In this fashion, all possible combinations and their outcomes can be deduced and then stored in the RB.

2.6 Defuzzification

As it was important to fuzzify data in order to apply non–deterministic approaches, it is important that we recover data from the output fuzzy consensus. The consensus is defuzzified to a crisp value within the universe of discourse of the output. Two defuzzification schemes have been considered to recall outputs from fuzzy consensus [52]: (1) The maximum–membership defuzzification scheme: This is the simplest scheme which allows for choosing the conclusion that has the maximal membership in the output fuzzy set:

$$\mu_c(\theta_c) = \sup\{\mu_c(\theta) \quad \theta \in \Theta\}, \tag{2.52}$$

where μ_c denotes the combined membership function $\mu : \quad \Theta \rightarrow [0,1]$ that assigns degrees of truth to the elements of Θ. If the universe of discourse Θ is a finite set of values $\{\theta_1, \theta_2, \cdots, \theta_n\}$, then one can replace the *supremum* with a maximum [52]:

$$\mu_c(\theta_c) = \max_{1 \leq i \leq n} \mu_c(\theta_{c_i}) . \tag{2.53}$$

(2) The fuzzy centroid defuzzification scheme: Using this scheme, one can directly compute the real–valued output as a convex combination of fit values. The fuzzy centroid $\bar{\theta}$ of the output variable $\theta \in \Theta$ is given by

$$\bar{\theta} = \frac{\int_{-\infty}^{\infty} \theta \, \mu(\theta) d\theta}{\int_{-\infty}^{\infty} \mu(\theta) d\theta} , \tag{2.54}$$

provided the integrals exist. In practice one can restrict fuzzy subsets to finite samples of the real line. Hence, one can replace the integrals over continuously infinite spaces with a ratio of discrete sums:

$$\bar{\theta} = \frac{\sum_{i=1}^{n} \theta_{c_i} \mu_c(\theta_{c_i})}{\sum_{i=1}^{n} \mu_c(\theta_{c_i})} . \tag{2.55}$$

The main advantage of using the fuzzy centroid is that it is unique and uses all the information resulting from output models.

Example: Suppose that the three edge maps shown in Fig 1.1 need to be fused. Assume that the surface normal edge map, the video edge map, and the range edge map are abbreviated by $\theta_1, \theta_2,$ and θ_3, respectively. Assume further that at a location x on these edge maps, the values of $\theta_1, \theta_2,$ and θ_3 are $0.62, 0.3,$ and 0.49, respectively. According to our data fusion system presented in sections 2.1 through 2.4, fusing $\theta_1, \theta_2,$ and θ_3 can be accomplished as follows:

Fuzzification: Let us assume that the universes of discourse $\Theta_1, \Theta_2,$ and Θ_3 are quantized equally into the three fuzzy values WE, MO, and ST. For this illustration, suppose

that the corresponding membership functions are assigned as

$$\theta_1 \longrightarrow [\mu_{we} = 0, \mu_{mo} = 0.43, \mu_{st} = 0.15],$$

$$\theta_2 \longrightarrow [\mu_{we} = 0.48, \mu_{mo} = 0.2, \mu_{st} = 0],$$

$$\theta_3 \longrightarrow [\mu_{we} = 0, \mu_{mo} = 0.96, \mu_{st} = 0].$$

Data Fusion (Rule Generation and Combination): Assume that there are four rules that result from the fuzzification scheme. These rules are

Rule 1: $\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(MO, WE, MO) \longrightarrow WE\}$ = an element of θ.

Rule 2: $\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(MO, MO, MO) \longrightarrow MO\}$ = an element of θ.

Rule 3: $\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, WE, MO) \longrightarrow MO\}$ = an element of θ.

Rule 4: $\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, MO, MO) \longrightarrow ST\}$ = an element of θ.

Figure 2.5 illustrates the above rules and their outputs. The strength of the support committed to the output θ of each rule depends on the strength of the membership values that are combined to make up that support. Using the fusion formula given by Eq. 2.20, we have

$$\mu^{(1)} = 0.5 + \left[\frac{(0.43 - 0.5)^3 + (0.48 - 0.5)^3 + (0.96 - 0.5)^3}{3}\right]^{1/3} = 0.81.$$

In a similar fashion, $\mu^{(2)}, \mu^{(3)}$, and $\mu^{(4)}$ are computed. The superscripts refer to the number of rules that are used for three given fuzzy inputs.

Defuzzification: To apply the centroid method for defuzzification, we compute the "center of gravity" of the fuzzy outputs. The centroid of the ith output fuzzy set is given in vector form as

$$\theta_{c_i} = (\theta_{c_1}, \theta_{c_2}, \theta_{c_3}) = (0.18, 0.5, 0.82).$$

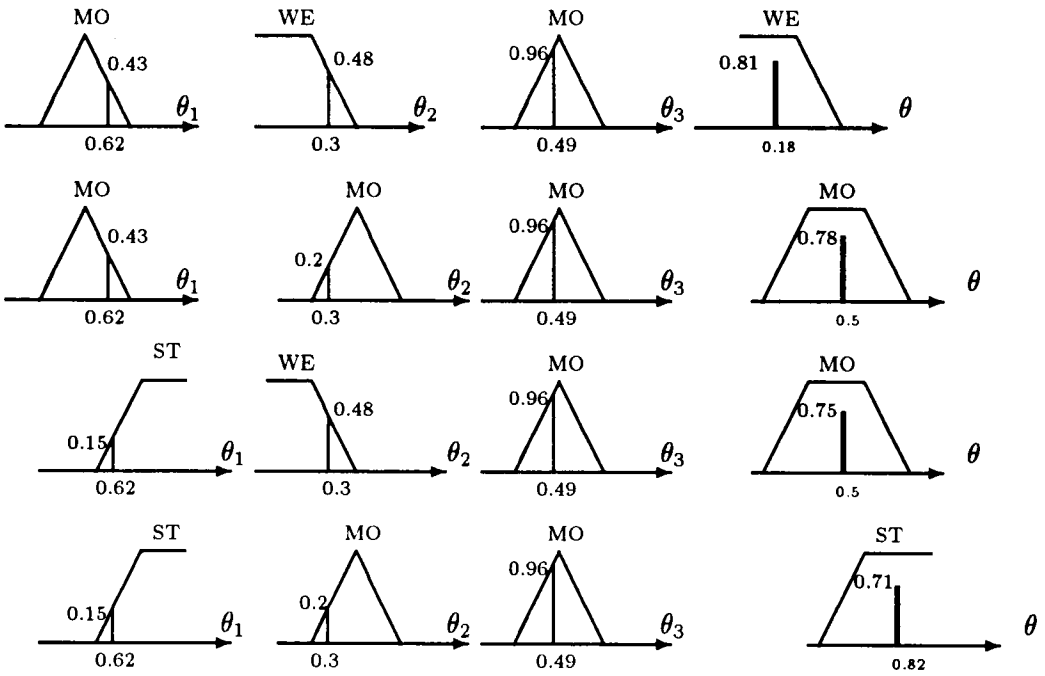


Figure 2.5: *Four outputs generated by four rules.*

The degree of support committed to each rule's output is defined as the "height" of the output fuzzy set. Hence, the height of the output fuzzy sets resulting from rule 1 through 4 are $0.81, 0.78, 0.75$, and 0.71, respectively. The cumulative output fuzzy set is

$$\overline{O} = (0.81, 1.53, 0.71),$$

where $\overline{O_1} = \mu^{(1)} = 0.81$, $\overline{O_2} = \mu^{(2)} + \mu^{(3)} = 1.53$, and $\overline{O_3} = \mu^{(4)} = 0.71$. $\overline{O_2}$ is computed as the sum of $\mu^{(2)}$ and $\mu^{(3)}$ because the two rules generate the same fuzzy value. Using Eq. 2.55, we have

$$\theta = \frac{0.18 \times 0.81 + 0.5 \times 1.53 + 0.82 \times 0.71}{0.81 + 1.53 + 0.71} = 0.49.$$

CHAPTER 3

The Use of Data Fusion for Segmentation: Experimental Results

In the literature, two approaches are often used for segmentation: (1) segmentation through edge detection and (2) segmentation through region growing. In this work, our approach to image segmentation is through edge detection. Before fusion, edges are extracted from different image modalities. An edge in an image corresponds to a discontinuity in the scene. Edge detection plays a central role in image segmentation. A scene can be partitioned into meaningful regions if one can successfully extract edges. There are many algorithms and procedures which are used to detect edges. The performance of a method to detect edges depends on the type of edge that needs to be extracted and the type of data that makes up this edge. Edge features are classified into three categories: (1) jump edges, (2) roof edges, and (3) texture edges. In practice, noise and digitization effects introduce difficulties in extracting complete boundaries of regions in an image. In the following, we briefly introduce several algorithms that have been used in the literature and discuss their performance. Our images are acquired from a real robot environment using an Odetics Laser Range Scanner. In order to achieve complete edge information from the scene, we apply our method, *i.e.*, the data fusion algorithm. Results from the use of the data fusion algorithm will be evaluated with respect to the presented edge detection techniques.

3.1 Edge Detection

Different approaches are taken by different authors to resolve the problem of edge detection. In this section, we will briefly introduce several edge detectors presented in the literature. We will implement them using real range and intensity images.

3.1.1 The Gradient Operator

The gradient [36] of an image function $f(x, y)$ is defined by

$$\mathbf{G}(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} G_x \\ \\ G_y \end{bmatrix} . \tag{3.1}$$

The magnitude of $\mathbf{G}(x, y)$ is given by

$$G(x, y) = \left[G_x^2 + G_y^2\right]^{\frac{1}{2}} . \tag{3.2}$$

The magnitude $G(x, y)$ is equal to the maximum rate of increase of $f(x, y)$ per unit distance in the direction of the gradient $G(x, y)$. Hence, at a discontinuity, the first derivative will give a peak response, which is an indication of "edginess." In contrast, the derivitive of a smooth region will give a low response, which is an indication of the absence of edginess. The direction of the gradient at a point (x, y) is given by

$$\alpha(x, y) = \arctan\left(\frac{G_x}{G_y}\right) . \tag{3.3}$$

An approach for computing the gradient components G_x and G_y is based on the use of the Sobel operator. Assume a 3×3 neighborhood labeled $x_1, x_2, \cdots, x_9$ around the point x_5. Using this mask, the components of the gradient at the center of the mask are given by

$$G_x(x, y) = (x_7 + 2 \times x_8 + x_9) - (x_1 + 2 \times x_2 + x_3) , \tag{3.4}$$

$$G_y(x,y) = (x_3 + 2 \times x_6 + x_9) - (x_1 + 2 \times x_4 + x_7)\,. \tag{3.5}$$

Computing G_x and G_y for each point in the image and substituting into Eq. 3.2 will result in a new image: the gradient image. Ideally, the gradient image highlights all edges and dims smooth regions.

3.1.2 The Laplacian Operator

The Laplacian operator is a second–order differential operator, which is given by

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\,. \tag{3.6}$$

Using a 3×3 neighborhood as the one described in computing the Sobel operator, the Laplacian can be approximated by

$$\nabla^2 f(x,y) = x_2 + x_4 + x_5 + x_7 - 4 \times x_5\,. \tag{3.7}$$

Computing $\nabla^2 f(x,y)$ at each point in the image $f(x,y)$,it will result in a new image: the Laplacian image. Ideally, a transition in the Laplacian image from negative to positive values or vice versa indicates the existence of discontinuity.

3.1.3 Zero–crossings (The Marr and Hildreth Approach)

Marr and Hildreth [61] presented a method of detecting edges of different intensities. Two ideas underlie their proposal: (1) A differential operator such as the Laplacian is used to detect discontinuities. (2) A Gaussian filter is used to smooth the original image. They emphasized the importance of using the Gaussian filter because it can be tuned to cope with different values of variations. Also, it has desirable properties for

implementation such as symmetry; a 1–D Gaussian distribution can be used in duplicate for the implementation of 2–D Gaussian distribution. The latter is given by

$$G(x,y) = \exp -\frac{x^2+y^2}{2\pi\sigma^2}\,. \tag{3.8}$$

If the image is given by the function $f(x,y)$, then the filtered image $I(x,y)$ is computed by convolving $f(x,y)$ with $G(x,y)$ as follows:

$$I(x,y) = G(x,y) * f(x,y)\,, \tag{3.9}$$

where $*$ denotes the convolution. Hence, edges are detected by localizing the zero–crossings of

$$\nabla^2 I(x,y) = \nabla^2[G(x,y) * f(x,y)]\,. \tag{3.10}$$

The mathematical identity allows us to move the ∇^2 operator inside the convolusion giving

$$\nabla^2 I(x,y) = [\nabla^2 G(x,y)] * f(x,y)\,. \tag{3.11}$$

3.1.4 Canny

Canny [11] defines an edge point as a local maxima of the operator G_n in some direction $\mathbf{n}$, where G is a 2–D Gaussian distribution

$$G(x,y) = \exp -\frac{x^2+y^2}{2\pi\sigma^2}\,, \tag{3.12}$$

and

$$G_n = \frac{\partial G}{\partial \mathbf{n}} = \mathbf{n}.\nabla G\,. \tag{3.13}$$

A good approximation of $\mathbf{n}$ is computed from the smoothed gradient direction [11]:

$$\mathbf{n} = \frac{\nabla(G * f)}{|\nabla(G * f)|}. \tag{3.14}$$

An edge point is defined to be a local maxima in the direction of $\mathbf{n}$; that is,

$$\frac{\partial}{\partial \mathbf{n}} G_n * I = 0. \tag{3.15}$$

From Eqs. 3.13 and 3.15, we obtain

$$\frac{\partial^2}{\partial n^2} G * I = 0. \tag{3.16}$$

So, the edge strength will be the magnitude of $G_n * f$. That is,

$$|G_n * f| = |\nabla[G * f]|. \tag{3.17}$$

There are strong links between the Marr–Hildreth and Canny operators. The main difference is that Canny's operator uses the directional gradient to give better localization of zero-crossings [11].

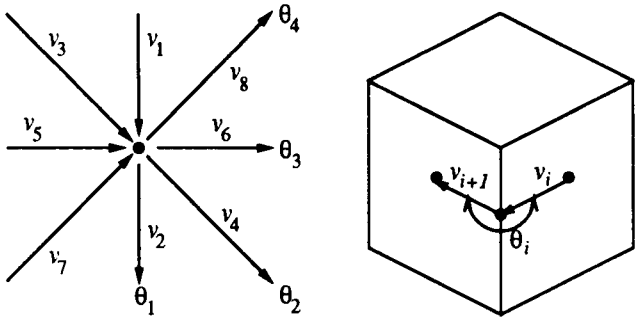
3.1.5 Angle of Curvature

Gil *et al.* [35] proposed the angle of curvature method to detect edges in range images. For each point in the image, eight neighboring vectors are calculated (Fig. 3.1). With the 3–D vectors $v_k, k = 1, \cdots, 8$, there are angles $\theta_m, m = 1, \cdots, 4$, which are computed as

$$\theta_1 = \arccos\left(\frac{v_1 . v_2}{\| v_1 \| \| v_2 \|}\right)$$

$$\theta_2 = \arccos\left(\frac{v_3 . v_4}{\| v_3 \| \| v_4 \|}\right)$$

$$\theta_3 = \arccos\left(\frac{v_5 . v_6}{\| v_5 \| \| v_6 \|}\right)$$



(a) (b)

Figure 3.1: *The angle of curvature: (a) vectors for computing angles θ_1–θ_4; (b) an illustration in 3–D.*

$$\theta_4 = \arccos\left(\frac{v_7 . v_8}{\| v_7 \| \| v_8 \|}\right)$$

To determine the angle of curvature A, the maximum θ_m is defined as follows:

$$A = \begin{cases} \max(\theta_1, \theta_3), & \text{for } \mid \theta_1 - \theta_3 \mid \geq \mid \theta_2 - \theta_4 \mid \\ \max(\theta_2, \theta_4), & \text{otherwise.} \end{cases} \tag{3.18}$$

The values of A range from 0 to 180. Large values of surface curvature indicate the existence of an edge. In this procedure, jump boundaries are assigned the maximum value of the surface curvature; *i.e.*, $A = 180$. Hence, significant changes in distances are equivalent to $A = 180$, where A is an unrefined measure of the surface curvature, and it is sensitive to noise in images. The authors recommend the use of a semi–thresholding procedure to enhance the appearance of the output image. For implementation of the above algorithms and procedures, real intensity and range images were acquired using the Odetics Laser Range Scanner. In addition, a normal image was computed from the range data. Canny's and Marr's operators are implemented using different values of standard deviation. All the operators mentioned in Section 3.1 are also implemented (see Figs. 3.2 through 3.4). Next, we will analyze these edge maps with respect to different types of edge features. Selected edge maps will be modeled and fused.

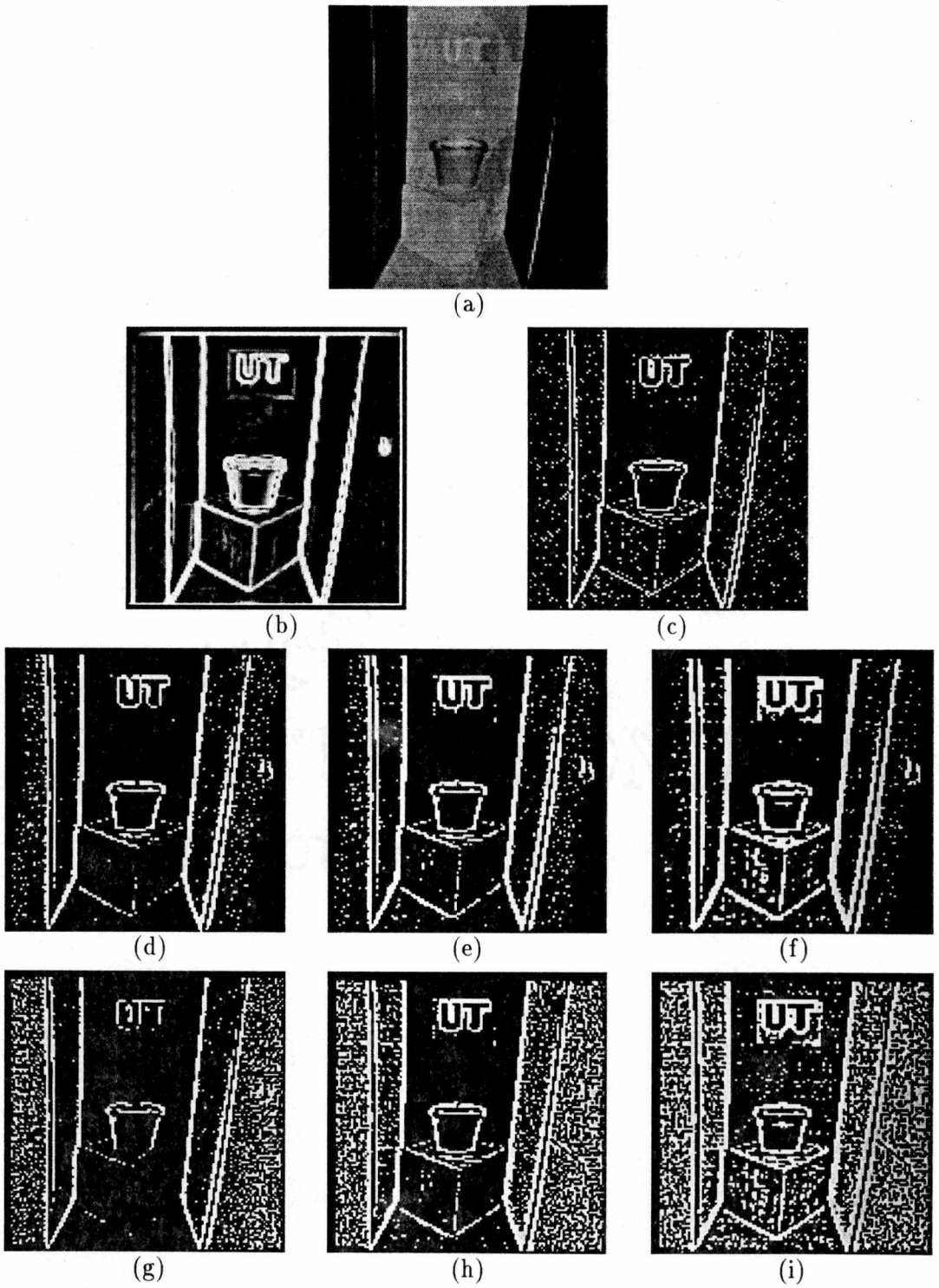


Figure 3.2: *Edge detection: (a) intensity image; (b) edges of (a) using the Sobel operator; (c) edges of (a) using the Laplacian operator; (d), (e), and (f) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (g), (h), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.*

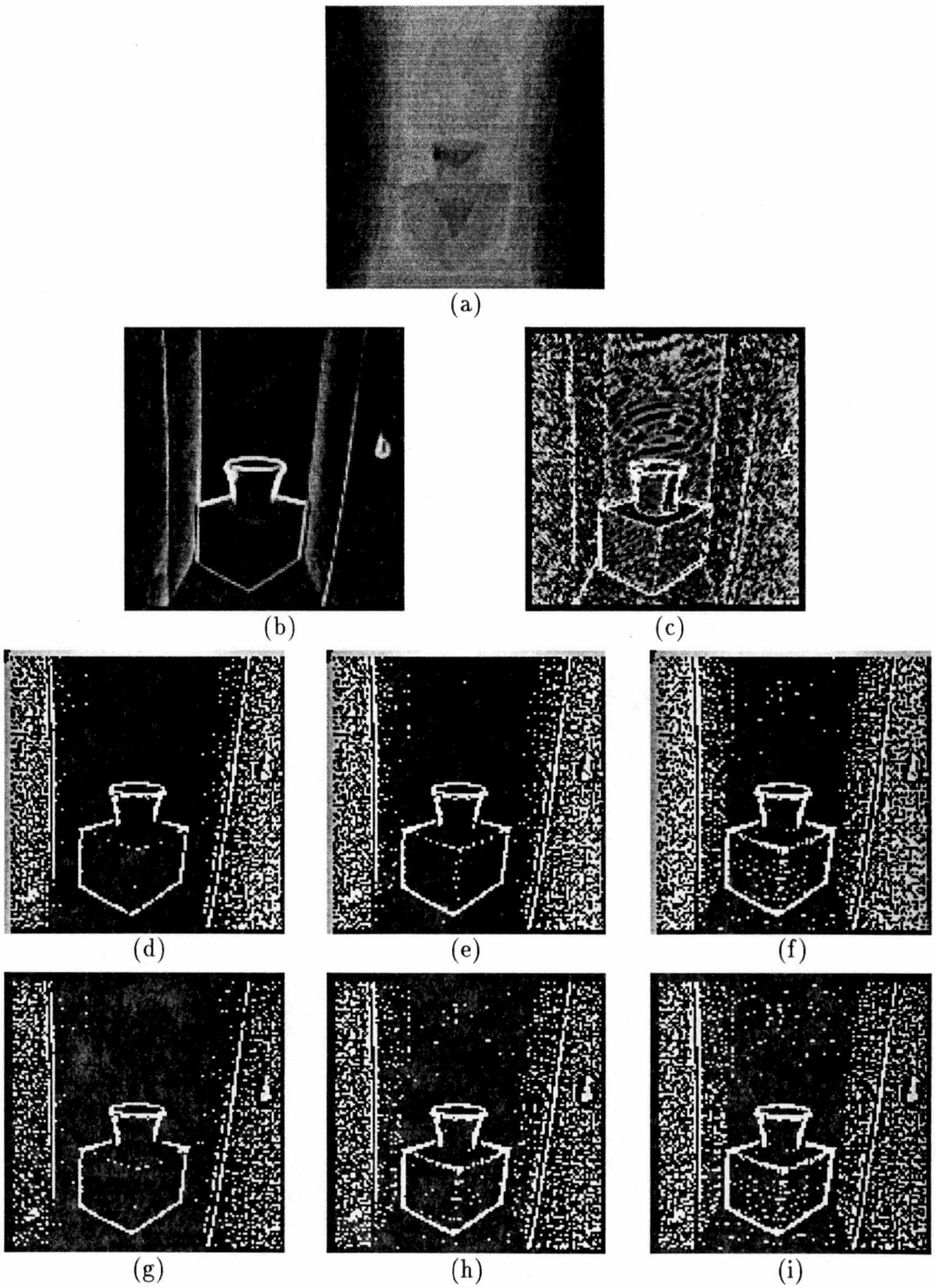


Figure 3.3: *Edge detection: (a) range image; (b) edges of (a) using the Sobel operator; (c) edges of (a) using the angle of curvature algorithm; (d), (e), and (f) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.93, 0.95$*, respectively; (g), (h), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.67, 0.69, 0.7$*, respectively.*

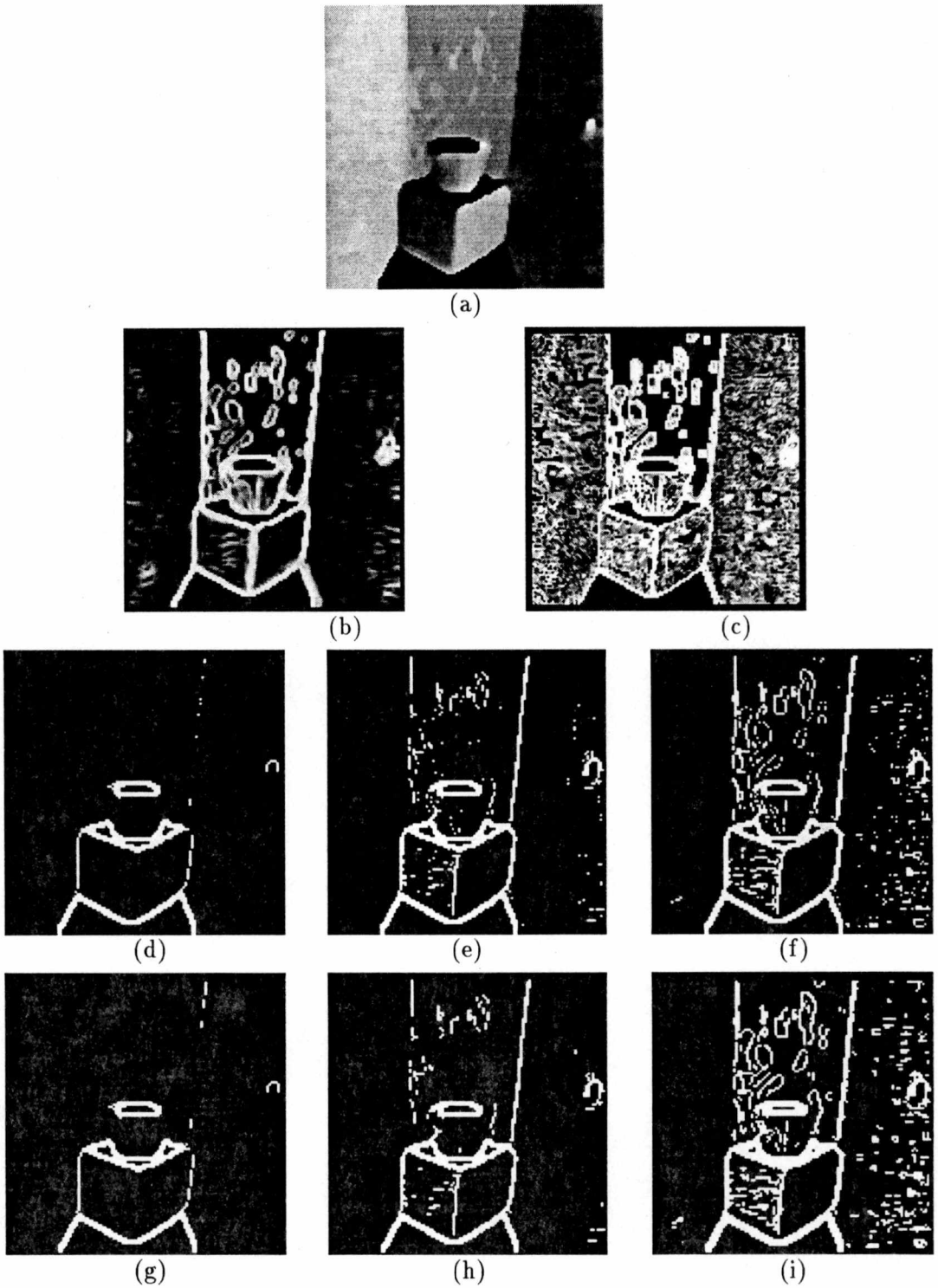


Figure 3.4: *Edge detection: (a) normal of a range image; (b) edges of (a) using the Sobel operator; (c) edges of (a) using the angle of curvature algorithm; (d), (e), and (f) edges of (a) using Canny's edge detector for* $\sigma = 0.7, 0.85, 0.9$, *respectively; (g), (h), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.5, 0.6, 0.7$, *respectively.*

3.2 Edge Features Classification

Edge features are classified into the three following categories: (1) jump edges, (2) roof edges, and (3) texture edges.

3.2.1 Jump Edges

In a range image, jump edges result from abrupt discontinuities in depth. In an intensity image, jump edges result from abrupt discontinuities in the gray level. Relatively, jump edges are easy to extract because a sudden change in depth of a range image or a sudden change in intensity of a reflectance image will give a peak in the first derivative or, equivalently, a zero–crossing in the second–order derivative.

By implementing the edge detection methods mentioned in the previous section, jump edges are extracted from the original range, intensity, and normal images. Both range and intensity images are reliable information sources for jump edges because variations in depth or intensity are strong enough so that the performance of most edge detectors is adequate. The normal image, however, does not generate complete edges because the background of the scene may have a surface normal that is identical to the surface of the small object (the cylinder which is on top of the box). As a result, these areas appear smooth with respect to the background. Hence, edge detectors are not able to identify most jump edges in the normal image. Figures 3.5 – 3.7 illustrate results regarding jump edges from intensity, range, and normal images, respectively. Also, an ideal edge detection picture is shown for comparison purposes.

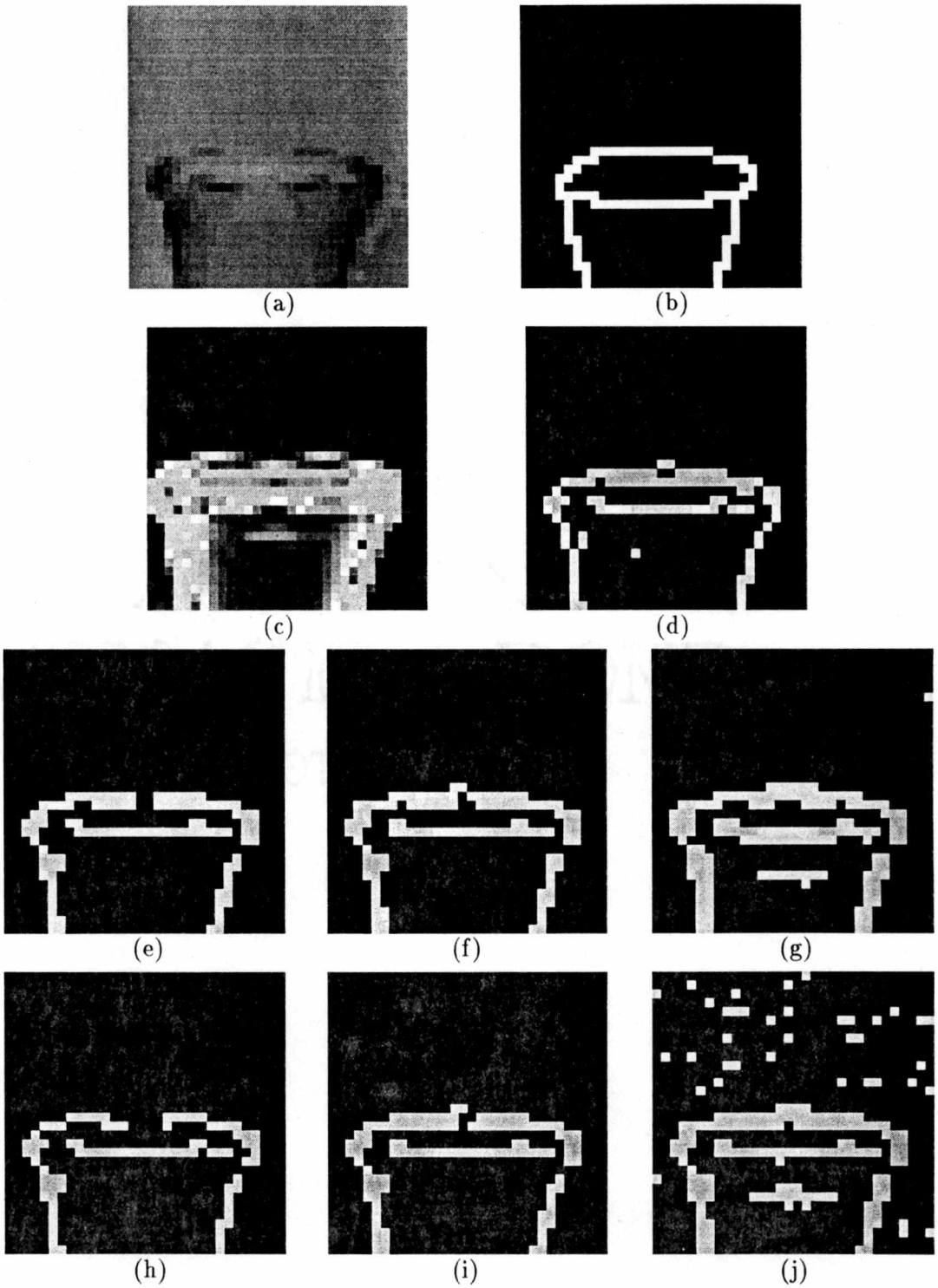


Figure 3.5: *Jump edges: (a) intensity image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.*

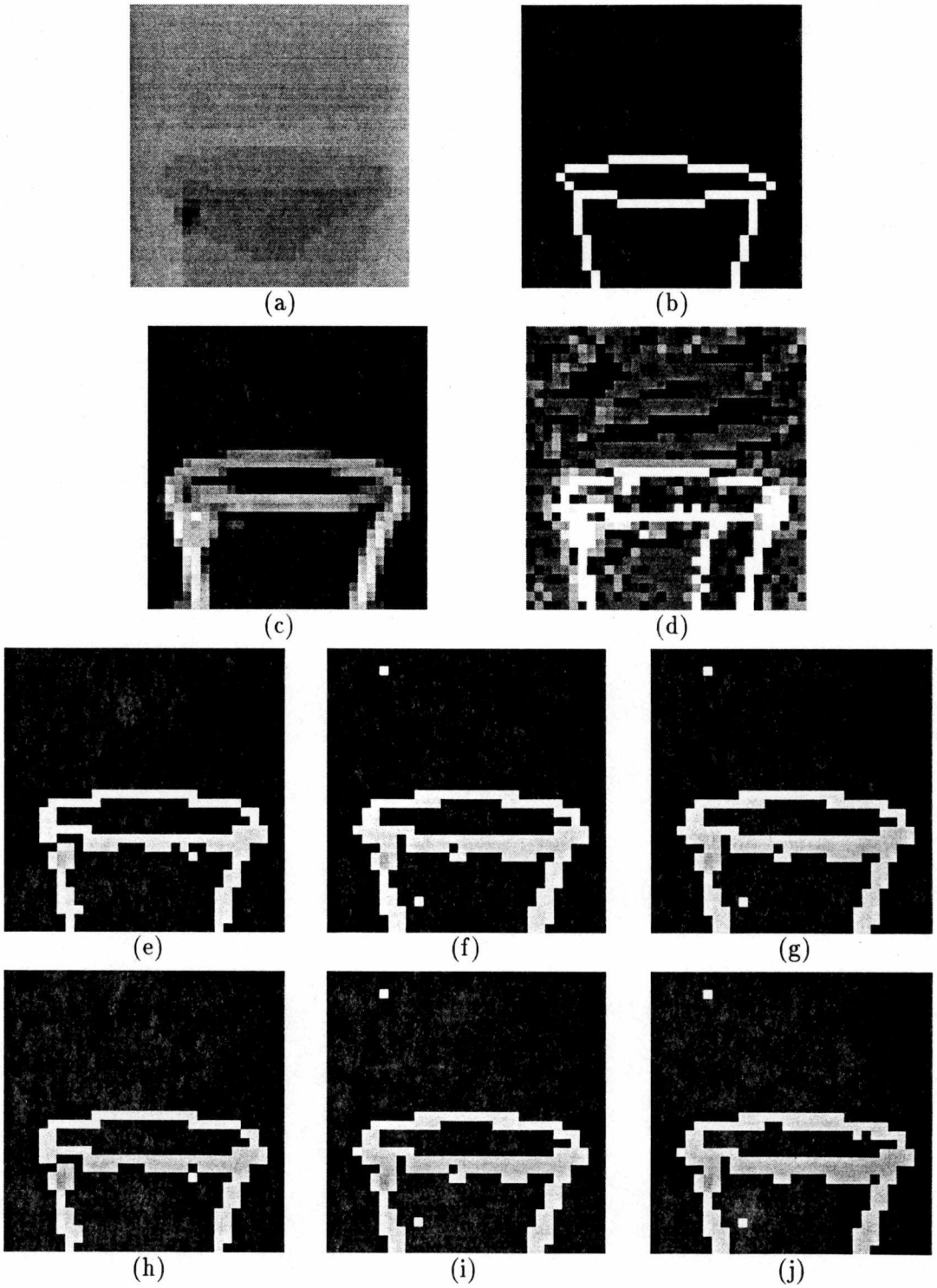


Figure 3.6: *Jump edges: (a) original range image; (b) ideal edge map; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.93, 0.95$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.67, 0.69, 0.7$, *respectively.*

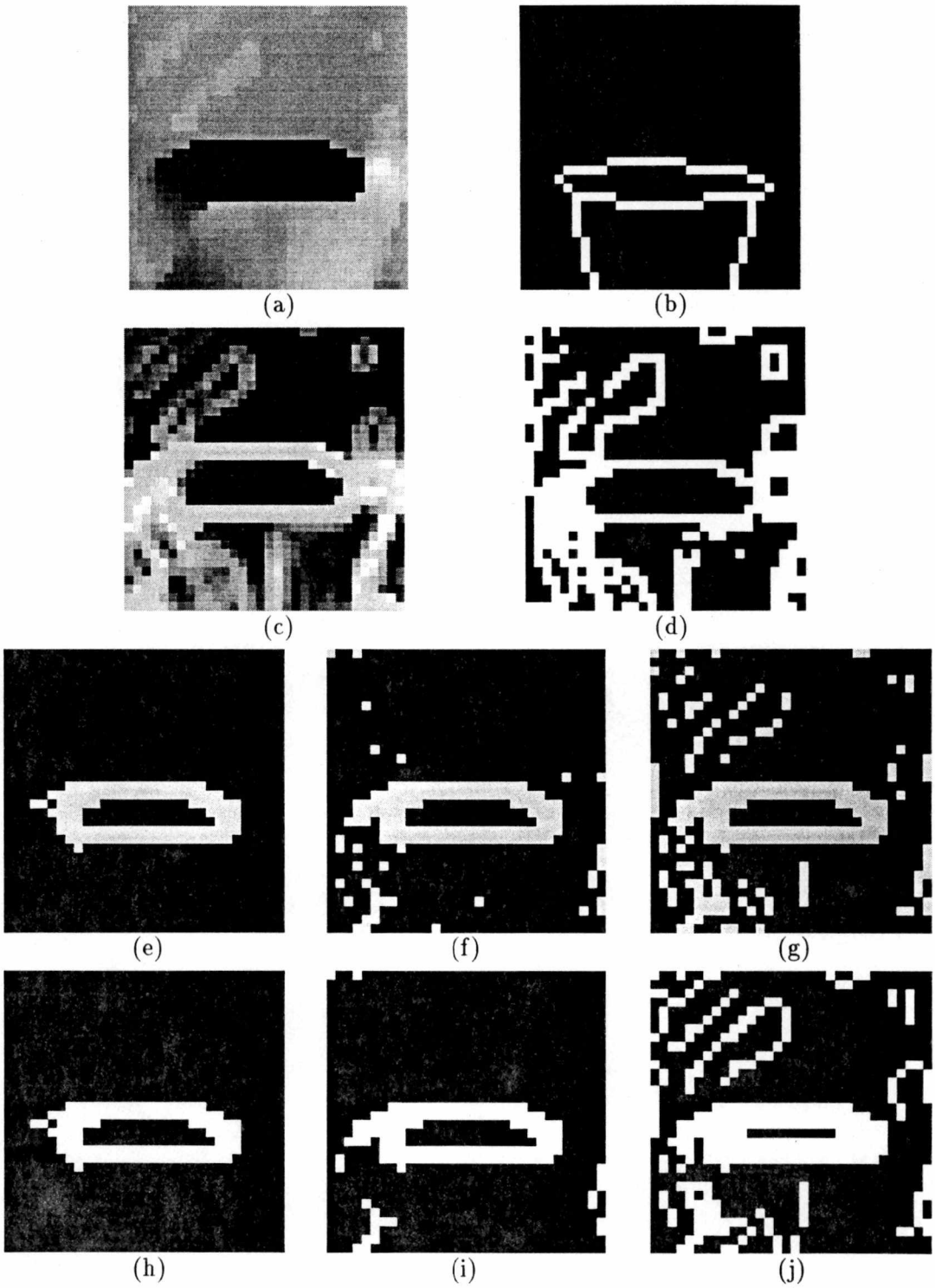


Figure 3.7: *Jump edges: (a) normal of a range image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.7, 0.85, 0.9$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.5, 0.6, 0.7$, *respectively.*

3.2.2 Roof Edges

This type of edge occurs at the intersection of surfaces with significantly different surface orientations. We will show that many standard edge detectors are not able to extract "reliable" roof edges from range images because there are no significant changes in depth values around these edges. Therefore, edge detection techniques that are based on differentiation often fail to detect small depth changes. Noise and digitization effects might introduce changes in depth which are of same order as those exhibited in roof edges. Figures 3.8 and 3.9 show the results of the roof edges from range and intensity images. Canny's and Marr's operators are tuned to recover roof edge information. There is a trade–off between recovering "complete" roof edge information and the increase of noise in the output image. An effective method for detecting this type of edge is to compute the surface normal of a range image [74]. Surface normal discontinuities occur at the intersection of surfaces in the original range image, *i.e.*, roof edges. Figure 3.10 shows complete roof edges from the surface normal image.

3.2.3 Texture Edges

Texture edges result from variations in the homogeneousness of surfaces. In this work, texture edges may result from artificial labels used to identify objects or artificial signs used as a title for the scanned environment. Texture edges cannot be extracted from range images because they do not introduce discontinuities in depth. However, intensity images are the most suitable sources of texture information. Figure 3.11 shows texture edges extracted from the intensity image. By comparing the results from intensity, range, and normal images, it is infeasible to depend on one type of image to provide a complete edge map for the scene. Figures 3.2 through 3.4 show incomplete edge

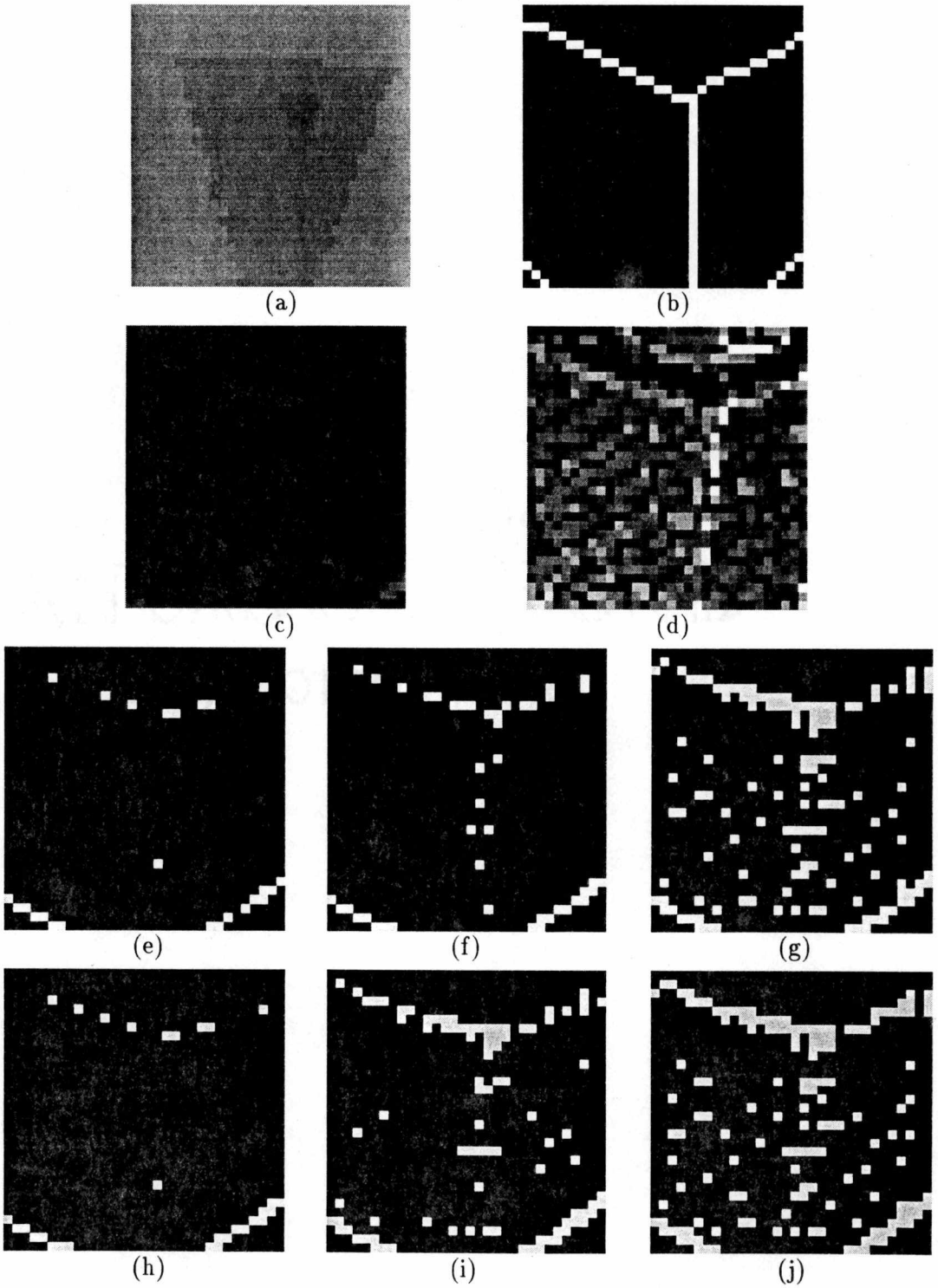


Figure 3.8: *Roof edges: (a) original range image; (b) ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.93, 0.95$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.67, 0.69, 0.7$, *respectively.*

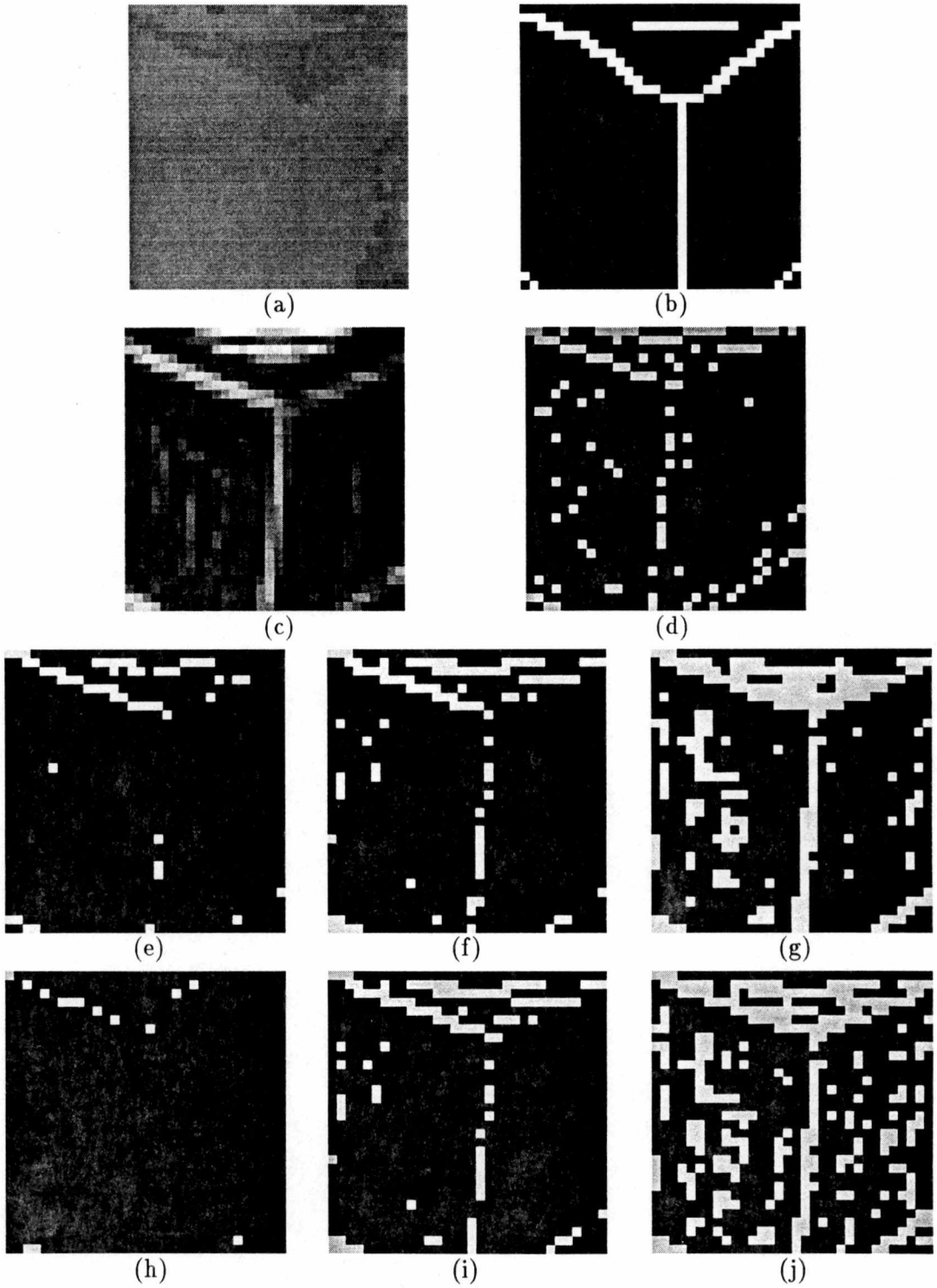


Figure 3.9: *Roof edges: (a) intensity image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the Laplacian operator; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.*

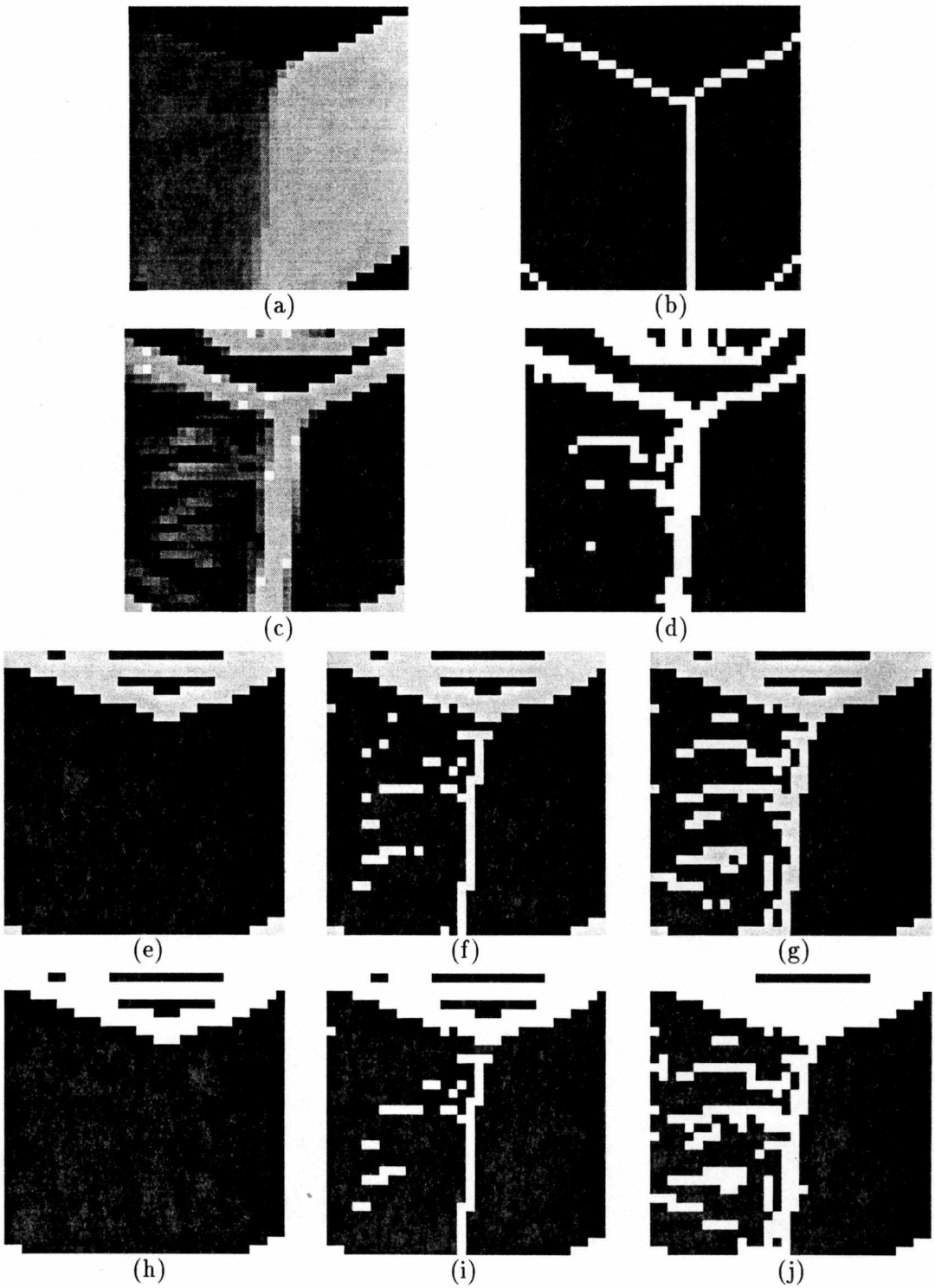


Figure 3.10: *Roof edges: (a) normal of a range image; (b) an ideal edge detection; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the angle of curvature algorithm; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.7, 0.85, 0.9$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.5, 0.6, 0.7$, *respectively.*

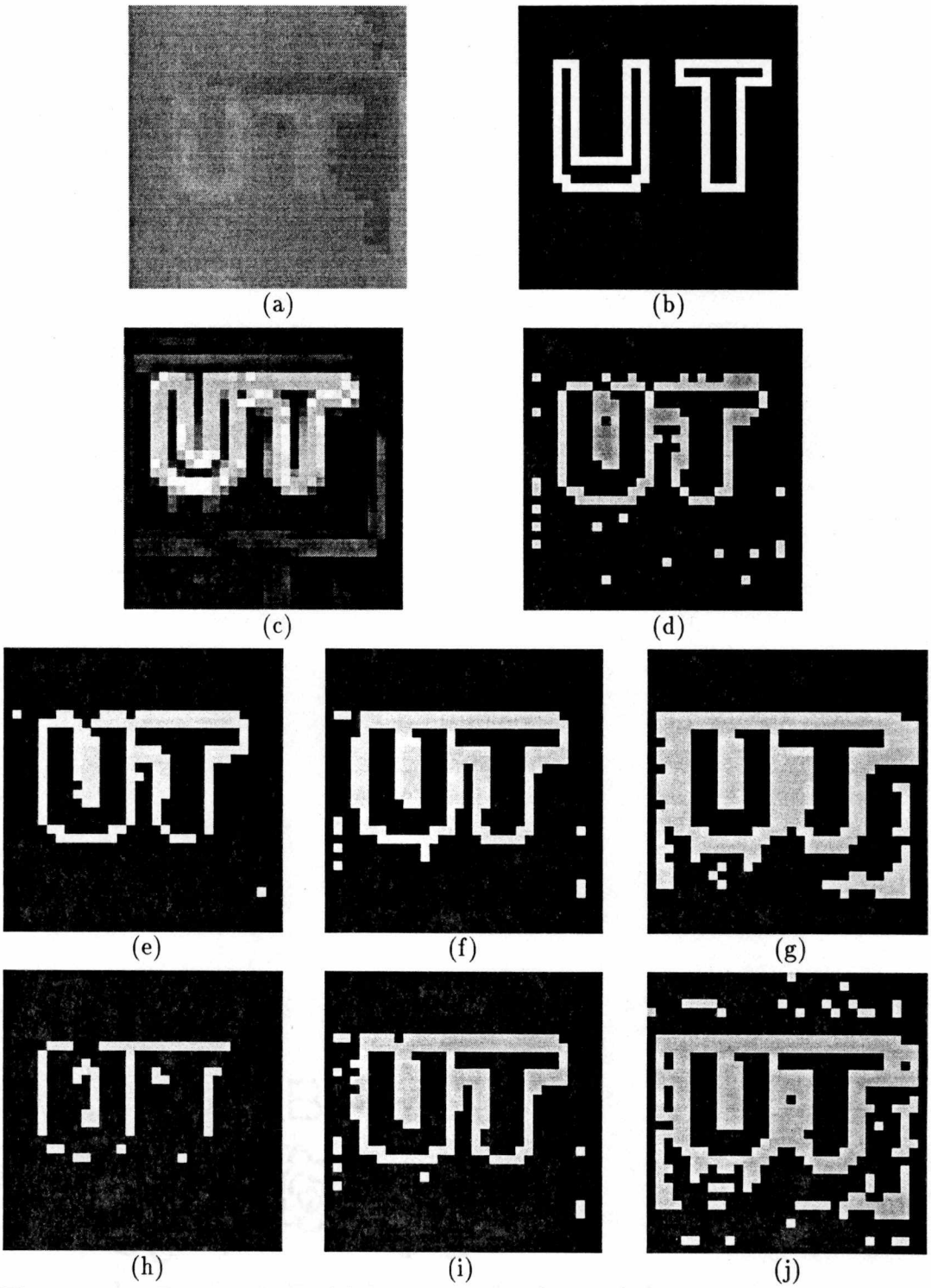


Figure 3.11: *Texture edges: (a) intensity image; (b) ideal edge map; (c) edges of (a) using the Sobel operator; (d) edges of (a) using the Laplacian operator; (e), (f), and (g) edges of (a) using Canny's edge detector for* $\sigma = 0.92, 0.98, 1.2$, *respectively; (h), (j), and (i) edges of (a) using Marr's edge detector for* $\sigma = 0.6, 0.65, 0.7$, *respectively.*

information in each image. However, we realize that some edge information may exist in one type of image and may not exist in another. The need for integrating information from several types of images becomes evident.

3.3 Data Fusion System Implementation

The implementation of our data fusion system is divided into three stages: (1) fuzzification of real data; (2) fusion; and (3) defuzzification. In this experiment, three edge maps are selected as inputs to the data fusion system. Our selection to the candidate inputs is based on the fact that we need to involve either two or three types of images in the fusion system.

3.3.1 Fuzzification

In this work, we examine the fusion of two images, then that of three images. In both cases, each type of image is represented by a universe of discourse, each of which is partitioned into several fuzzy regions. Fuzzy values are, therefore, overlapped within each universe of discourse. In Chapter 2, we discussed the possibility of partitioning universes of discourse into three, five, or seven overlapped fuzzy values (subsets). It is common to have an odd number of fuzzy values. The problem which we are dealing with is the combination of several edge maps. To fuzzify the edginess of an edge map, we assign the variable θ to measure the strength of the edge map at hand. The fuzzification of edginess is tied to our thinking about an edge. The interpretation of a given edge is greatly influenced by our experience and common sense in image segmentation. We can characterize an edge by its strength. Because the domain of an edginess variable can

have positive values on a gray scale, we can normalize the edginess' strength within the interval $[0, 1]$. We start out with a minimum number of overlapped fuzzy values. The assigned fuzzy values are (1) weak edge (WE), (2) moderate edge (MO), and (3) strong edge (ST). These fuzzy values characterize weak, moderate, and strong portions of the universe of discourse (edginess). The percentage of overlapping among these clusters of edginess strength will be determined experimentally. When selecting the number of fuzzy values, the sensitivity of data and the simplicity of the design are considered. In this design, we assume that three fuzzy values are enough to adequately characterize these edge maps. However, we will increase the number of fuzzy values to five for each edge map and conduct a comparison between the two modeling schemes. If the use of five or more fuzzy values introduces no significant improvement, then the use of three fuzzy values is preferred because a small number of fuzzy quantization universes of discourse yields a simpler data fusion system. That is, the design of knowledge bases is simpler, and then the real time of execution is shorter.

There are two types of fuzzy values, continuous and discrete. Both types are shown in Figs. 2.1 and 2.4. The discrete type is represented by triangular and trapezoidal shapes, which are compatible with the continuous type. The discrete type simplifies computations. Figure 3.12 shows membership function graphs for the fuzzy subsets of four universes of discourse. The values $\theta_1, \theta_2, \theta_3$, and θ represent the strength of edginess of the normal, video, range, and the fused edge maps, respectively.

3.3.2 Fusion

Two cases of fusion are considered:

(1) Fusion of two images: An edge map of an intensity image using Canny's edge

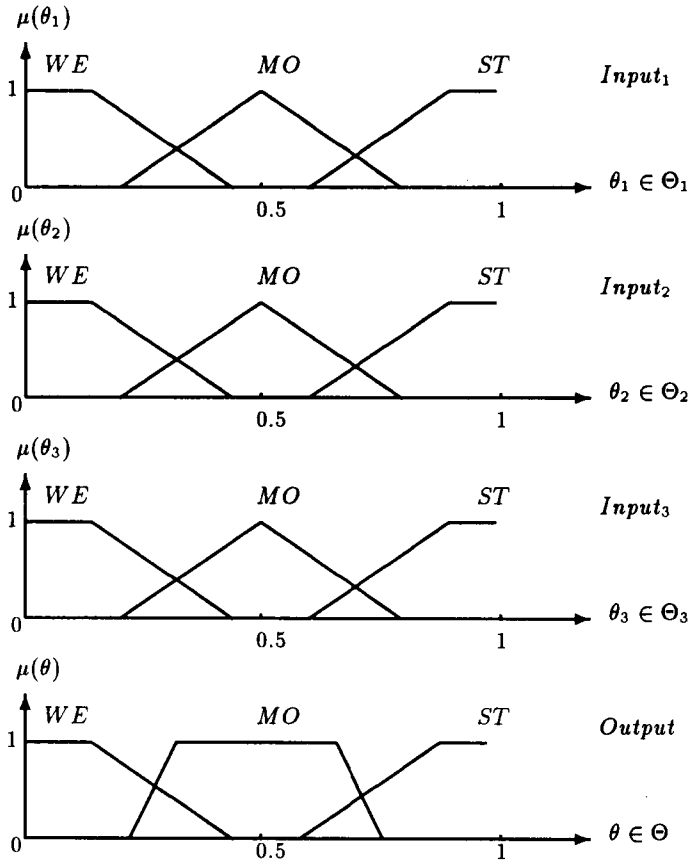


Figure 3.12: *Four membership functions for three inputs and one output.*

detector (see Fig. 3.3–(d)) and an edge map of the normal of a range image using the Sobel operator (see Fig. 3.4–(b)) are considered for fusion. These two edge maps can complement one another to provide a better edge map from either inputs. The normal image gives complete roof edges, but it cannot provide information about texture edges. The intensity image gives information about texture edges, but it cannot provide complete information about roof edges. Variations in intensity at the roof edges are not strong enough to be detected by most edge detectors. Let $\Theta_1, \Theta_2,$ and Θ represent the universes of discourse of edges from the intensity image, the normal image, and the

fused image, respectively. Two subtasks are carried out:

(1) Rule generation: To construct a complete RB, we examine the two input edge maps. For each location in the scene, there will be two assessments reported by these two inputs. In each possible combination, we determine a single variable that suggests an action. For instance,

$$\theta_1 \oplus \theta_2 = \{(MO, ST) \longrightarrow ST\} = \text{an element of } \theta.$$

We index the rows using the three fuzzy values that partition the edge map from the intensity image. This input is represented by θ_1. We index the columns using the three fuzzy values that partition the edge map from the surface normal of the range image. This input is represented by θ_2. We design these rules by anticipating possible input combinations of θ_1 and θ_2 and by applying common sense and our experience in image segmentation to determine the combined edge map θ.

Because roof edges are ST in the normal edge map, the fused edge map must be ST or at least MO.

$$\theta_1 \oplus \theta_2 = \{(ST, WE) \longrightarrow MO\} = \text{an element of } \theta.$$

$$\theta_1 \oplus \theta_2 = \{(ST, MO) \longrightarrow ST\} = \text{an element of } \theta.$$

$$\theta_1 \oplus \theta_2 = \{(ST, ST) \longrightarrow ST\} = \text{an element of } \theta.$$

Because texture edges are WE, MO, or ST in the video edge map and are WE in the normal edge map, the fused edge map must be assigned the values as the video edge map:

$$\theta_1 \oplus \theta_2 = \{(WE, WE) \longrightarrow WE\} = \text{an element of } \theta.$$

$$\theta_1 \oplus \theta_2 = \{(WE, MO) \longrightarrow MO\} = \text{an element of } \theta.$$

θ_1 \ θ_2	WE	MO	ST
WE	WE	MO	ST
MO	MO	MO	ST
ST	MO	ST	ST

Figure 3.13: *A rule base matrix: θ_1 and θ_2 represent the normal and video edge maps, respectively.*

$$\theta_1 \oplus \theta_2 = \{(WE, ST) \longrightarrow ST\} = \text{an element of } \theta.$$

Because both the normal edge map and the video edge map provide support for jump edges, the fused edge map must be assigned values that are in relative agreement with both edge maps. There are 9 possible combinations altogether. The generated rules must be enough to construct a complete inference system for managing uncertainty. The RB of 9 rules is represented by a 3×3 linguistic matrix, which is illustrated in Fig. 3.13.

(2) Support combination: In each rule, the outcome of a fuzzy rule is supported to a degree $\mu(\theta)$. This cumulative degree is computed by combining the degrees of support provided by all input fuzzy values that make up this rule. The support is determined using the fusion formula:

$$\mu_i(\theta) = e + \left[\frac{(\mu_i(\theta_1) - e)^3 + (\mu_i(\theta_2) - e)^3}{2}\right]^{1/3},$$

where $i = 1, 2, \cdots, 9$. The number of rules used at each given input depends on the overlapping status between fuzzy values.

(2) Fusion of three images: In addition to the same pair of images that were used in case (1), an edge map from the range image is used (see Fig. 3.2–(e)). Normally, range images show strong jump boundaries. Let $\Theta_1, \Theta_2, \Theta_3,$ and Θ represent the universes

of discourse of edges from the range, the video, the normal, and the fused edge maps, respectively. Similar to case (1), two subtasks are performed:

(1) Rule generation: We index the rows using the three fuzzy values that partition the edge map from the intensity image. This input is represented by θ_2. We index the columns using the three fuzzy values that partition the edge map from the surface normal of the range image. This input is represented by θ_3. To include the edge map from the range image, we fix the fuzzy variable θ_1 as WE, MO, and ST. We construct the RB following the same common sense that we used in case (1). However, including an edge map from a range image would provide us with reliable information about jump edges. Therefore, whenever θ_1 is ST or MO, it is considered a reliable indication for the existence of jump edges regardless of the values of θ_2. For instance,

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, WE, ST) \longrightarrow ST\} = \text{an element of } \theta.$$

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, WE, MO) \longrightarrow ST\} = \text{an element of } \theta.$$

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, WE, WE) \longrightarrow MO\} = \text{an element of } \theta.$$

There are 27 possible combinations altogether. The RB of 27 rules is represented by three 3×3 linguistic matrices. Figure 3.14 shows the rule bases that are designed to recover edges from three images.

(2) Support combination: In each rule, the outcome of a fuzzy rule is supported to a degree $\mu(\theta)$. The cumulative output of each rule is determined using the fusion formula:

$$\mu_i(\theta) = e + \left[\frac{(\mu_i(\theta_1) - e)^3 + (\mu_i(\theta_2) - e)^3 + (\mu_i(\theta_3) - e)^3}{3}\right]^{1/3},$$

where $i = 1, 2, \cdots, 27$.

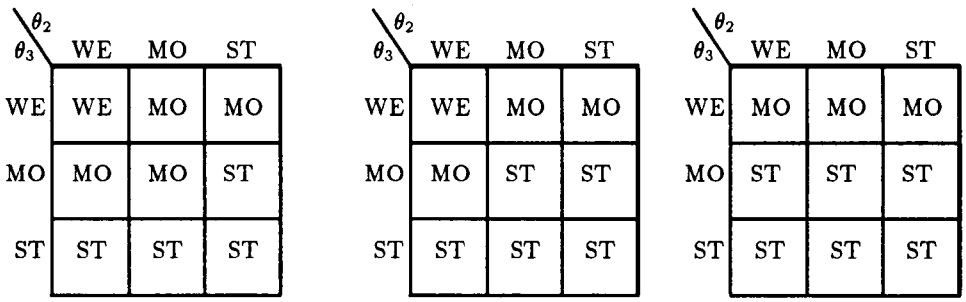


Figure 3.14: *Rule base matrices: (a) $\theta_1 \equiv WE$; (b) $\theta_1 \equiv MO$; (c) $\theta_1 \equiv ST$.*

3.3.3 Defuzzification

In this section, we defuzzify the output of the fusion system in order to recover crisp output values. The defuzzification scheme is based on the centroid method. At each value of the fuzzy output, the crisp value is computed as follows:

$$\theta = \frac{\sum_{i=1}^{n} \mu_i \mu_c(\theta_{c_i})}{\sum_{i=1}^{n} \mu_i},$$

where n is the number of used rules at a given input and $\mu_c(\theta_i)$ is the centroid of the fuzzy output at the ith rule. The results of this fuzzy logic–based data fusion are illustrated in Fig. 3.15.

3.4 Image Segmentation

Image segmentation is the process of dividing an image into regions, each of which corresponds to a "homogeneous" surface in a scene. Our approach has been based on extracting closed contours around surfaces. If complete edge information can be extracted, a reliable segmentation process can therefore be achieved. Edge maps without the use of data fusion are often broken. Consequently, open contours lead to the merger

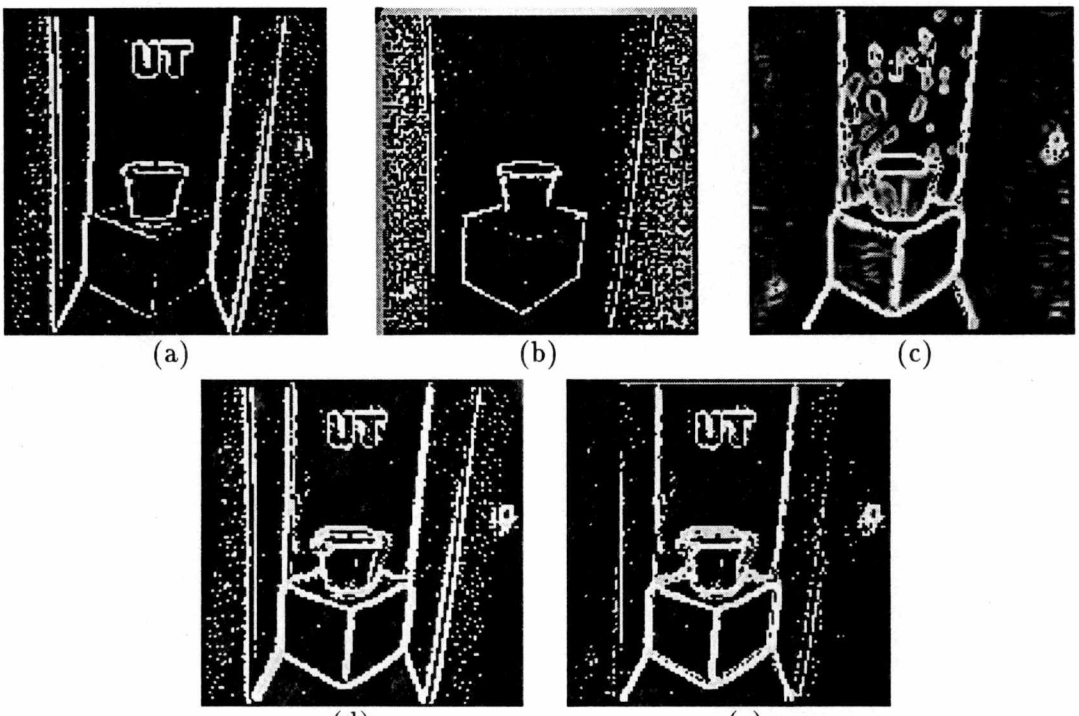


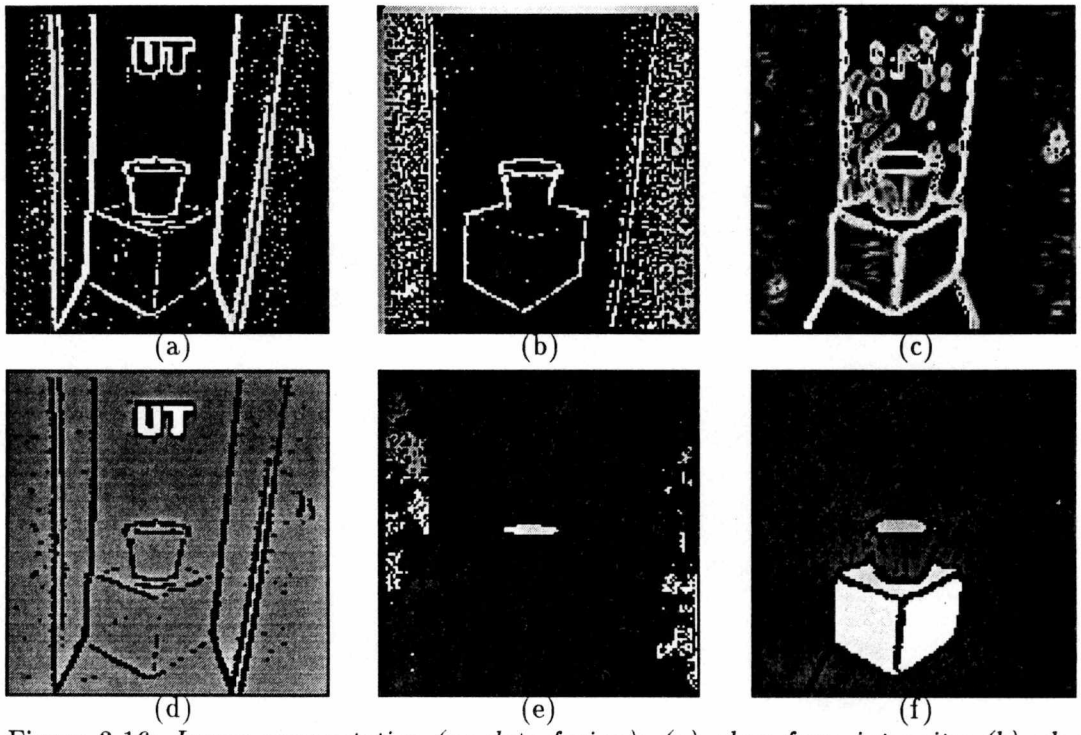
Figure 3.15: *Data fusion: (a) edges from video; (b) edges from range; (c) edges from surface normal; (d) Case (1): fused edges of (a) and (c); (e) Case (2): fused edges of (a), (b), and (c).*

of two or more different regions as if they belonged to the same surface. In contrast, fusing edge maps from different types of images closes most broken edges and allows the identification of the major regions of the image. Experimentally, we search for regions that are contained within closed contours. These regions are then isolated and labeled as distinct segmented surfaces. In Fig. 3.16, segmentation and labeling results of the images are processed individually. In this, the surface normal edges allow us to extract seven different regions. The segmentation of the intensity edges and the range edges have been less effective to extract regions. The segmented range and surface normal images are not able to identify regions contained within texture edges. Likewise, the segmented video image fails to identify most major regions in the image. Fusing two

or three images improved the results of the segmentation, usually resulting in more regions being identified. Figure 3.17 shows segmentation results based on the fusion of only two images. Figure 3.18 shows segmentation results obtained from the fusion of three edge maps. In both segmentation cases, the results show significant improvements over the segmentation of separate maps. However, the use of three edge maps has given better results than that of two edge maps. In the former and by inspection, all major regions are isolated and labeled. When comparing the two segmentation cases, the use of three edge maps is closer to an ideal segmentation case than to that of two edge maps; the ideal case will be shown in Section 3.6. However, the use of three edge maps caused larger data bases and longer real time for execution. Two different pictures have also been fused and segmented in Figs. 3.19 and 3.20. Results will systematically be evaluated. Furthermore, we will implement other methods for data fusion, such as the SBA and Dempster's rule of combination. The same experimental evaluation will be conducted.

3.5 Experimental Results Using Other Data Fusion Approaches

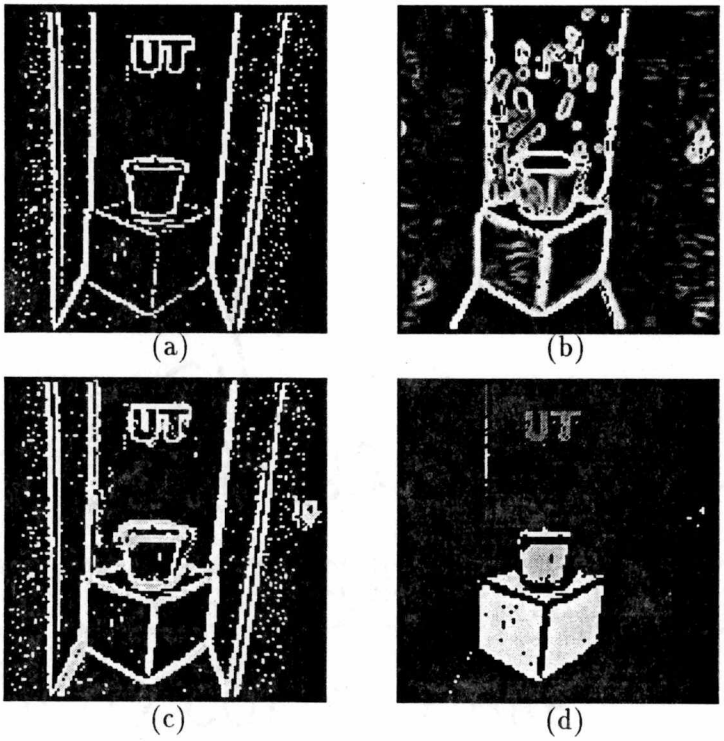
The Super Bayesian Approach (SBA) and Dempster's rule of combination are two of the most popular methods for data fusion. For the sake of comparison, we will implement these two methods. In order to achieve a meaningful comparison, we use the same images that were used in our approach. Since Dempster's rule of combination only allows a pairwise orthogonal sum computation for combination, we will only fuse two images at a time. We will, therefore, compare results obtained from our approach and those results obtained from the SBA and Dempster's rule of combination.



(a) (b) (c)

(d) (e) (f)

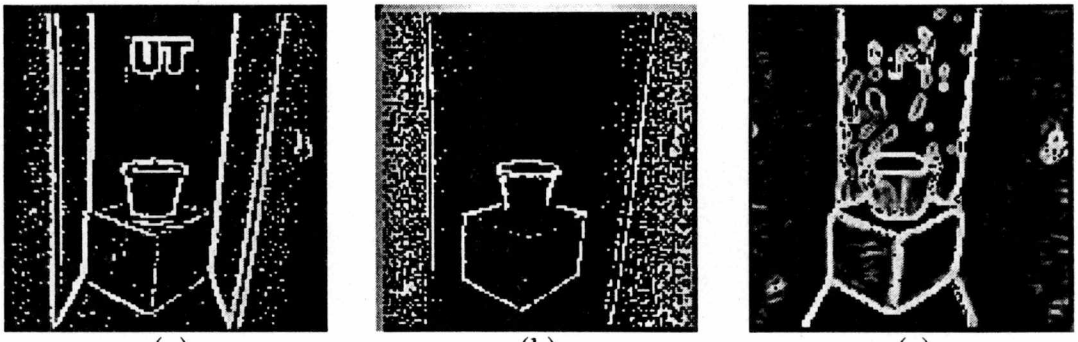
Figure 3.16: *Image segmentation (no data fusion): (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d), (e), and (f) segmented images of (a), (b), and (c), respectively.*



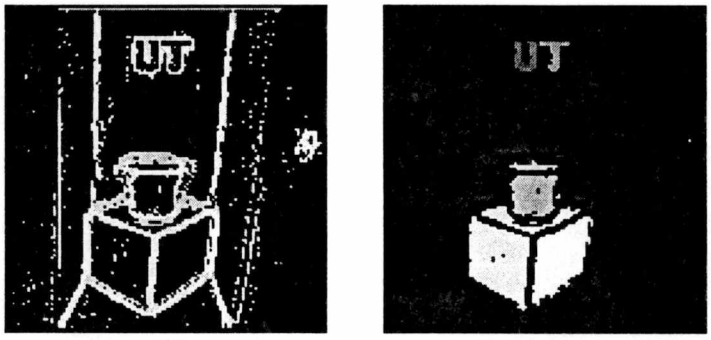
(a) (b)

(c) (d)

Figure 3.17: *Image segmentation (after data fusion): (a) edges from intensity; (b) edges from surface normal; (c) fused edges from (a) and (b); (d) segmented image of (c).*



(a) (b) (c)



(d) (e)

Figure 3.18: *Image segmentation (after data fusion): (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d) fused edges from (a), (b), and (c); (e) segmented image of (d).*

3.5.1 The Super Bayesian Approach

The SBA is a direct application of Bayes' theorem. A complete derivation of the SBA is shown in Appendix D. Let us consider the use of the SBA to combine two edge maps. We scale the strength of edges in the feature maps to be within the unit interval $[0, 1]$. If we consider one the feature maps as a decision maker (*a priori* probability) and the other feature map as a likelihood probability, the combined support is the *a posteriori* probability, which is computed by the SBA as follows:

$$S(\theta) = \frac{S_1(\theta)S_2(\theta)}{S_1(\theta)S_2(\theta) + (1 - S_1(\theta))(1 - S_2(\theta))}$$

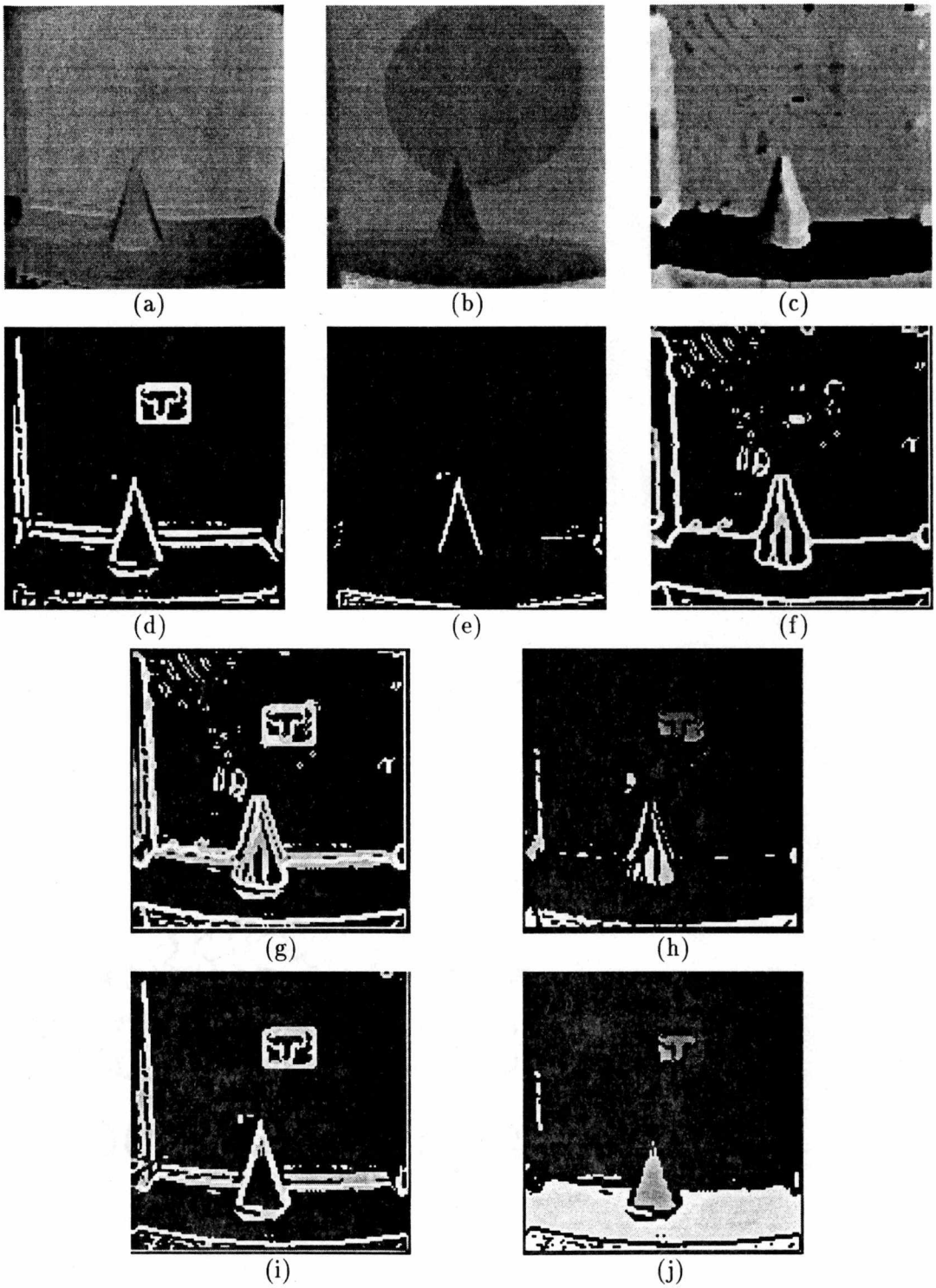


Figure 3.19: *Data fusion: (a) intensity; (b) range; (c) surface normal; (d), (e), and (f) edges from (a), (b), and (c), respectively; (g) fused edges from (d) and (f); (h) segmented image of (g); (i) fused edges from (d), (e), and (f); (j) segmented image of (i).*

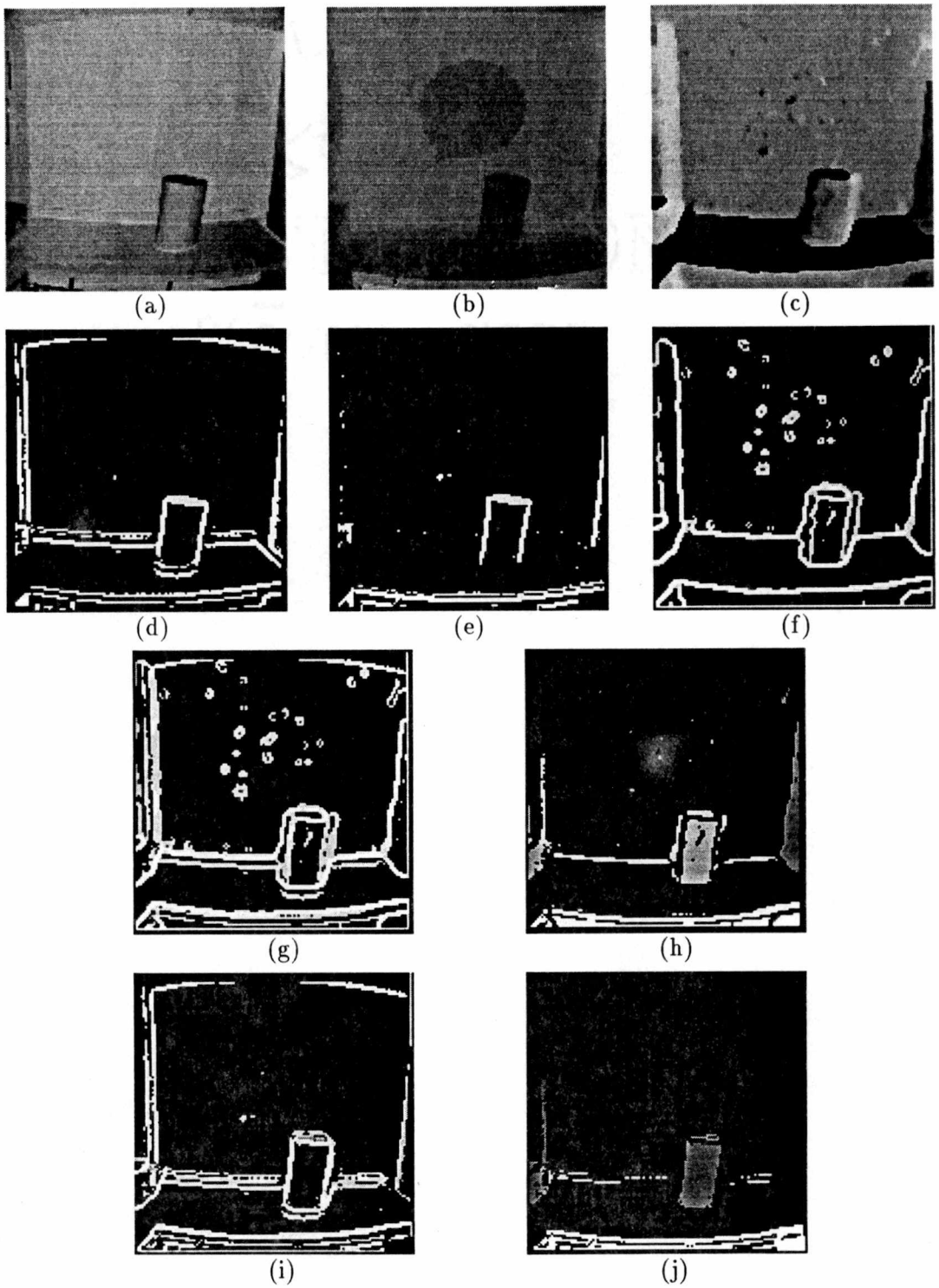


Figure 3.20: *Data fusion: (a) intensity; (b) range; (c) surface normal; (d), (e), and (f) edges from (a), (b), and (c), respectively; (g) fused edges from (d) and (f); (h) segmented image of (g); (i) fused edges from (d), (e), and (f); (j) segmented image of (i).*

Our interpretation of the combined support using the SBA is as follows: (1) the event θ can be interpreted as "there is an edge;" (2) $S_1(\theta)$ and $S_2(\theta)$ are probabilities committed by S_1 and S_2 in support of θ, respectively; and (3) $S(\theta)$ is the combined support of θ. After fusing $S_1(\theta)$ and $S_2(\theta)$, the values of $S(\theta)$ are rescaled to represent the strength of an edge map. This method is used to combine feature maps from the surface normal of a range image and an intensity image, both of which describe the same scene. Results of data fusion and segmentation using the SBA are given in Fig. 3.21.

3.5.2 Dempster's Rule of Combination

Let $S_1(\theta)$ and $S_2(\theta)$ be two basic probability numbers obtained from two information sources, and let $\theta \in \Theta$ be an event or a focal point that is supported by both S_1 and S_2. If both sources of information are in support of the same event $\theta \in \Theta$, then the combined support $S(\theta)$ is given by Bernoulli's rule of combination, which is a special case of Dempster's rule of combination (see Appendix D):

$$S(\theta) = 1 - (1 - S_1(\theta))(1 - S_2(\theta)).$$

Let θ be interpreted as "there is an edge." If the strengths of edges in two edge maps are normalized between 0 and 1, S_1 and S_2 can be interpreted as two basic probability numbers supporting the event θ. The combined support $S(\theta)$ is then rescaled to represent the combined edge map. This method is used to combine edge maps from the surface normal of the range image and the intensity image. The combined edge map is used to obtain a segmented image output for the scene. Results from Dempster's rule of combination are given in Fig. 3.22.

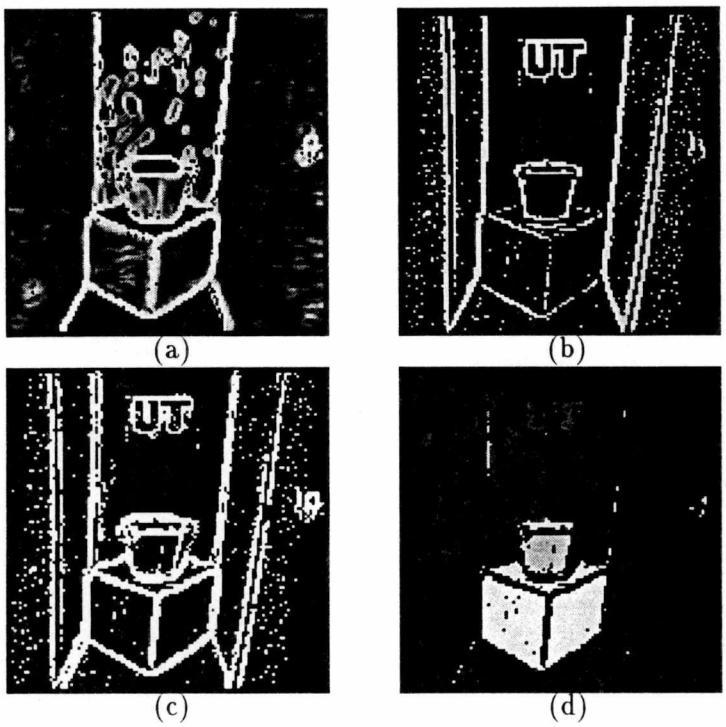


Figure 3.21: *Image segmentation using the SBA: (a) edges from surface normal; (b) edges from intensity; (c) fused edges from (a) and (b); (d) the segmented image of (c) (11 extracted region).*

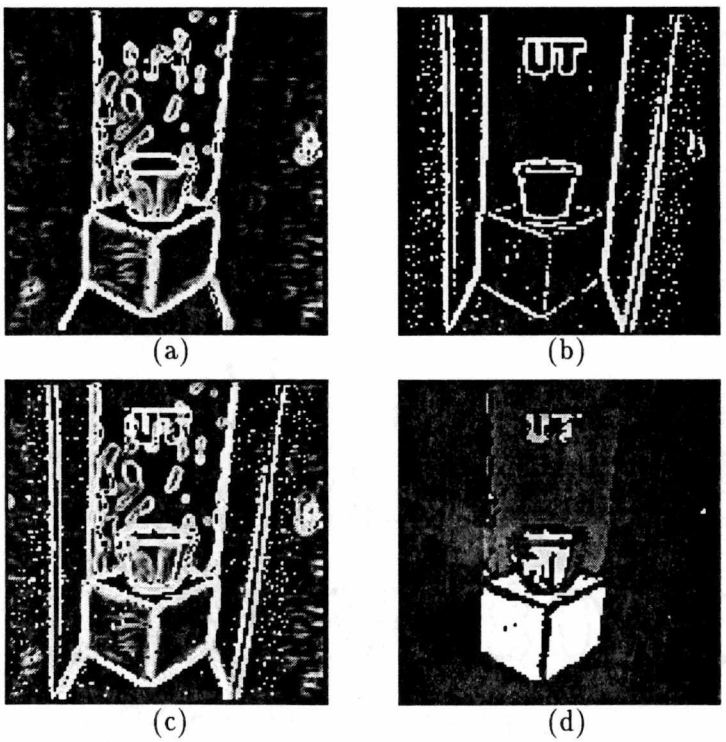


Figure 3.22: *Image segmentation using Dempster's rule of combination: (a) edges from surface normal; (b) edges from intensity; (c) Fused edges from (a) and (b); (d) the segmented image of (c) (16 extracted regions).*

3.6 Experimental Evaluation

In order to evaluate the segmentation results, we test these results against an ideal segmented image which is synthetically generated (see Fig. 3.23–(a)). Our quantitative evaluation criterion is based on the degree of similarity between the actual and the ideal segmented images. This degree of similarity may be quantified by computing a metric distance between the two segmented images. A segmented image is viewed as a point in n–dimensional space formed by the segmented image's attributes. Because a segmented image contains isolated regions, the attributes of the point are defined as the number of pixels in each region. Let two segmented images be represented by two vectors of observations $X' = [x_1, x_2, \cdots, x_n]$ and $Y' = [y_1, y_2, \cdots, y_n]$, where x_i and y_i, $i = 1, 2, \cdots, n$, denote the number of pixels in the *ith* region in the images, and the prime ($'$) represents transposition. A degree of similarity can be achieved by computing the *Euclidean* distance between X and Y as follows:

$$d_{xy} = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{1/2} . \tag{3.19}$$

The smaller d_{xy} is, the greater the degree of similarity between x and y is. Since d_{xy} is a metric distance, we normalize it with respect to the worst case, *i.e.*, the maximum distance between the worst case and the ideal case. Then, we quantify the measure of similarity, s_{xy}, in percent as follows:

$$s_{xy} = (1 - \frac{d_{xy}}{d_{xw}})100\% , \tag{3.20}$$

where d_{xw} is the *Euclidean* distance between an ideal segmented image , X, and the worst possible case which may introduce a maximum *Euclidean* distance. The latter may arise when there is no region extracted in the segmented image and, therefore, the

Table 3.1: *Results of segmentation through different data fusion approaches.*

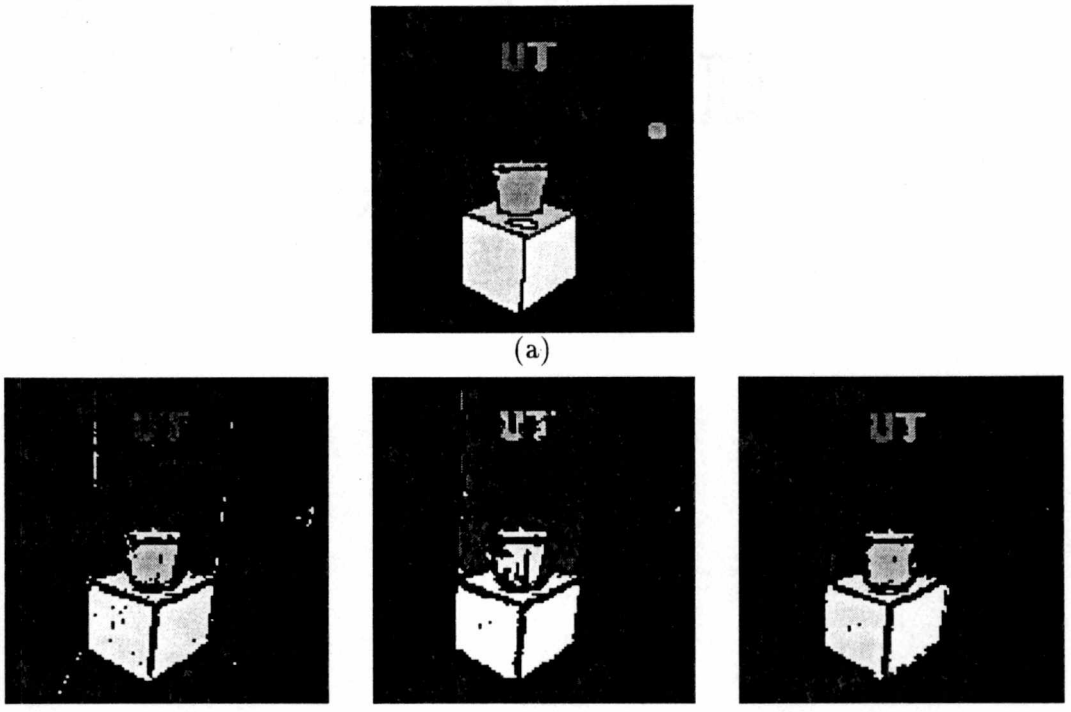
Regions	Ideal case I	Worst case W	Bayesian B	Dempster D	Fusion F
1	664	1878	620	570	586
2	500	0	466	467	472
3	189	0	138	138	117
4	12	0	0	0	3
5	286	0	221	132	240
6	38	0	59	38	36
7	45	0	3	3	1
8	79	0	73	53	76
9	65	0	59	36	62
10	0	0	12	0	3
11	0	0	5	10	0
12	0	0	3	5	0

whole image is considered as one region. Applying Eqs. 3.19 and 3.20, we can compute the degree of similarity between two segmentation cases. In the case where the vectors have different dimensions, the smaller vector is expanded by *augmenting* it by zeros. From each approach, its performance is determined by evaluating the similarity of the output of the approach to the ideal case. The extracted regions and their contents are given in Table 3.1. The degrees of similarity between the ideal case and those obtained from the SBA, Dempster's rule, and the fusion formula are (1) $s_{ib} = 90\%$, (2) $s_{id} = 85\%$, and (3) $s_{if} = 91\%$, respectively. We have numerically quantified the performances of our approach, the SBA, and Dempster's rule of combination. The results obtained from our approach and the SBA are the most similar ones to those obtained from the ideal case because some of the desirable properties are absent from Depmster's rule of combination. Specifically, Dempster's rule of combination conducts no test for certainty and is unable to solve conflict. In the following, we examine the effects of these properties on data fusion through the use of real data.

Test for Certainty: Three 3×3 windows are fused using Demspter's rule of combi-

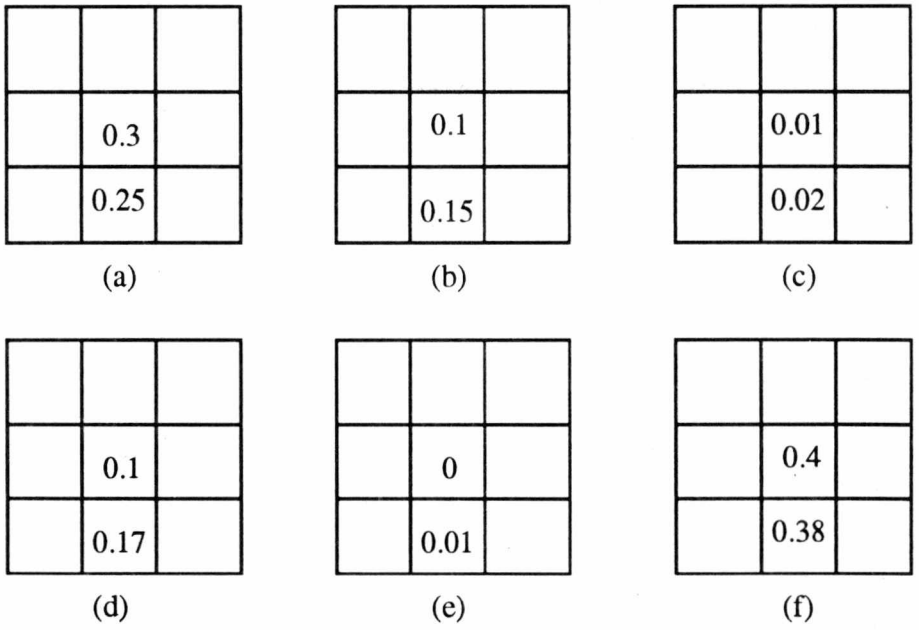
nation, the SBA, and the new fusion formula. Values of these small regions resulted from a modeling scheme for the three edge maps that were used in the previous experimentation. These regions and the output of the fusion are shown in Fig. 3.24. The use of Demspter's rule of combination may always result in overestimation of the given values. After fusion, noise transformed to the combined image with greater support. In contrast, the SBA and the new fusion formula have confirmed the weaknesses of these given assessments; the combined assessment is either underestimated or minimized.

Conflict Resolution (Convexity): Another example is given in Fig. 3.25. The use of Dempster's rule of combination resulted in support of one of the sources of information, and the other was ignored. The SBA and the new fusion formula solved the conflict between these two assessments by conducting a convex combination. The SBA and the fusion formula have favored 0.1 over 0.8. That is, 0.8 is considered to be less certain than 0.1. In both methods, the midpoint of the support region is considered to be the most uncertain piece of support.



(a) (b) (c) (d)

Figure 3.23: *Image segmentation: (a) ideal case; (b) SBA; (c) Dempster's rule; (d) the fusion formula.*



(a) (b) (c)

(d) (e) (f)

Figure 3.24: *Data fusion: (a), (b), and (c) regions to be fused; (d) using the fusion formula; (e) using the SBA; (f) using Dempster's rule.*

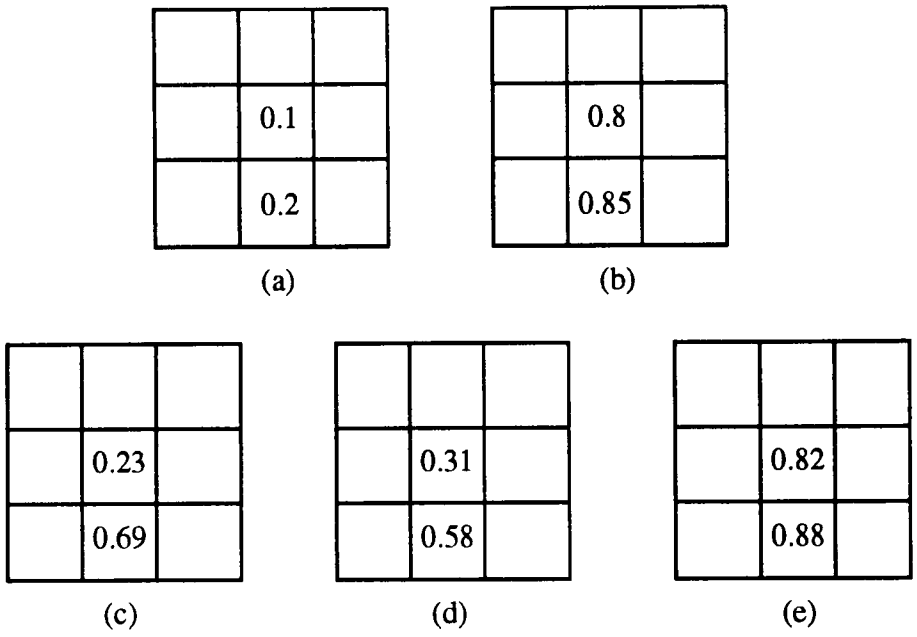


Figure 3.25: *Data fusion: (a) and (b) regions to be fused; (c) using the fusion formula; (d) using the SBA; (e) using Dempster's rule.*

CHAPTER 4

Effects of Fusion Parameters on Segmentation

We have conducted a large number of experiments on a variety of different images. Significant improvement in image segmentation was obtained using the data fusion algorithms. During the course of segmentation, three parameters were fixed at constant values: the standard deviation (noise control) of the Gaussian filter used in Canny's and Marr's edge detectors, the combination weighting parameter of the fusion formula, and the number of overlapping fuzzy values resulting from the fuzzification scheme. By varying these parameters, we would like to study their effects on the segmentation process.

4.1 Effects of Noise Level

Although Canny's and Marr's operators are able to produce reasonable edge feature maps, it should be noted that they cannot be automated. In both algorithms, a Gaussian filter is used to smooth out images before localizing edges. Selecting a standard deviation for the Gaussian filter is done by trial–and–error. As a consequence, we have found that there is a trade–off between extracting weak edges, the most difficult ones in range images, and the level of noise in the edge feature maps. In this section, we would like to consider various values of the standard deviation, which in turn introduces different levels of noise in the edge maps. Noisy edge maps will be fused using the SBA, Dempster's rule of combination, and the new fusion formula. Segmentation

Table 4.1: *Results of segmentation after adding noise ($\sigma = 0.8$).*

Regions	Ideal case I	SBA B	Dempster D	Fusion F
1	664	538	558	564
2	500	480	452	443
3	189	94	109	108
4	12	5	0	3
5	286	198	190	230
6	38	35	36	36
7	45	8	3	0
8	79	71	70	75
9	65	56	55	61
10	0	88	37	0
11	0	100	6	0
12	0	15	6	0
13	0	26	0	0
14	0	7	0	0

results obtained from each approach will be evaluated against the ideal segmentation case used previously. Two cases of noisy images are considered through the use of two different values of the standard deviation. Figures 4.1 and 4.2 show *noisy* edge maps, the segmentation results from the ideal case, the SBA, Dempster's rule, and the new fusion formula for $\sigma = 0.8$ and $\sigma = 1.2$, respectively. In the following section, noise tolerance of the approaches will be presented.

4.1.1 Experimental Evaluation

Adding noise to one of the input edge maps affected the segmentation results obtained from the data fusion approaches. The actual and ideal segmentation results are given in Tables 4.1 and 4.2. Using $\sigma = 0.8$ yields degrees of similarity between the ideal case and those obtained from other approaches as follows:

(1) $s_{ib} = 83\%$

(2) $s_{id} = 80\%$

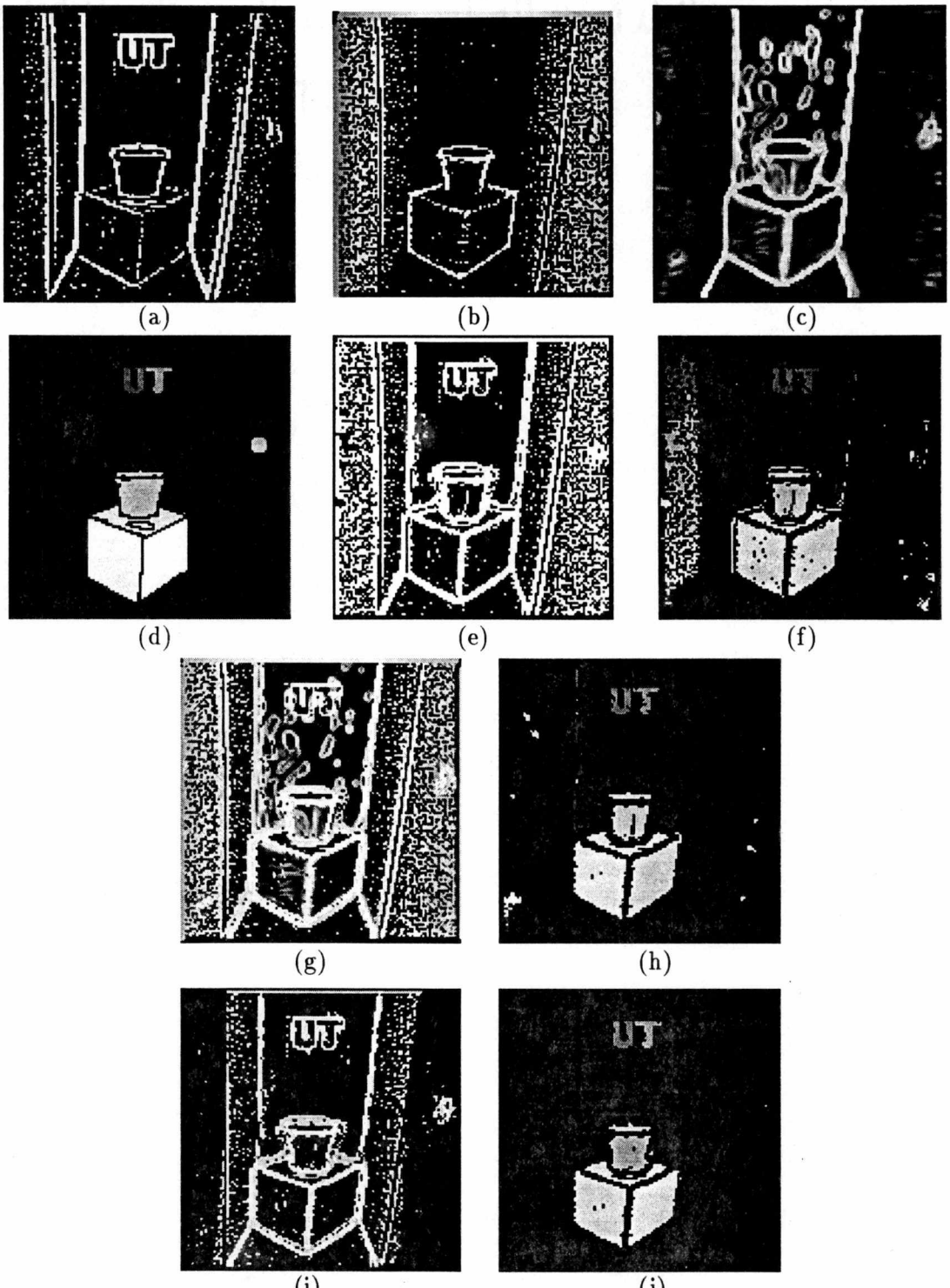


Figure 4.1: *Data fusion: (a), (b), and (c) edges from intensity, range, and surface normal with noise added ($\sigma = 0.8$), respectively; (d) ideal segmentation; (e) fused edges using the SBA; (f) segmented image of (e); (g) fused edges using Dempster's rule; (h) segmented image of (g); (i) fused edges using the fusion formula; (j) segmented image of (i).*

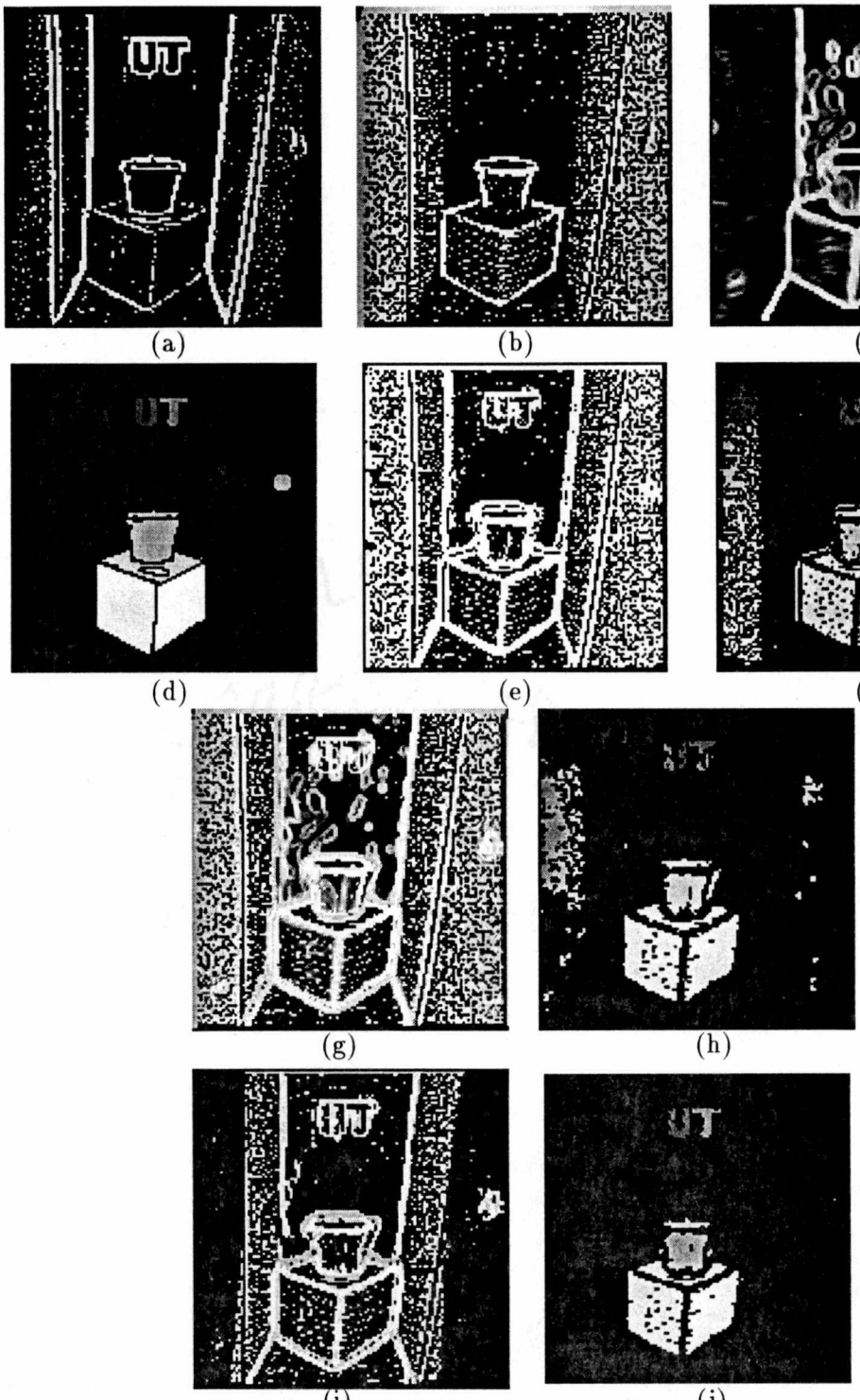


Figure 4.2: *Data fusion: (a), (b), and (c) edges from intensity, range, and surface normal with noise added ($\sigma = 1.2$), respectively; (d) ideal segmentation; (e) fused edges using the SBA; (f) segmented image of (e); (g) fused edges using Dempster's rule; (h) segmented image of (g); (i) fused edges using the fusion formula; (j) segmented image of (i).*

Table 4.2: *Results of segmentation after adding noise ($\sigma = 1.2$).*

Regions	Ideal case I	SBA B	Dempster D	Fusion F
1	664	469	503	514
2	500	407	410	407
3	189	89	106	102
4	12	5	0	3
5	286	180	188	212
6	38	35	37	36
7	45	4	0	0
8	79	41	45	65
9	65	53	43	58
10	0	40	0	0
11	0	230	20	0
12	0	69	408	0
13	0	30	39	0
14	0	25	9	0

(3) $s_{if} = 88\%$.

Using $\sigma = 1.2$ increases the level of noise in one of the input images. In this case, the degrees of similarity between the ideal segmentation results and those obtained from other approaches are as follows:

(1) $s_{ib} = 73\%$

(2) $s_{id} = 65\%$

(3) $s_{if} = 84\%$.

Table 4.3 shows the effect of three levels of noise on the segmentation results obtained from the SBA, Dempster's rule, and the new fusion formula. Glancing through Table 4.3, we notice that the segmentation results using the fusion formula are closer to the ideal segmentation results than are those obtained from the SBA or from Dempster's rule of combination. Further, results obtained from the SBA or from Dempster's rule of combination are significantly influenced by the added noise to one of the input edge maps. The added noise causes the degrees of similarity to drop from 90% to 73% and

Table 4.3: *Summary of noise effects on segmentation.*

$\sigma = 0.7$	$\sigma = 0.8$	$\sigma = 1.2$
$s_{ib} = 90\%$	$s_{ib} = 83\%$	$s_{ib} = 73\%$
$s_{id} = 85\%$	$s_{id} = 80\%$	$s_{id} = 65\%$
$s_{if} = 91\%$	$s_{if} = 88\%$	$s_{if} = 84\%$

from 85% to 65% in the SBA and Dempster's rule, respectively. In contrast, the added noise has less effect on results obtained from the new fusion formula where the degree of similarity dropped only from 91% to 84%. This observation shows the notion that the new fusion formula is more robust than the SBA or Dempster's rule of combination. In this work, robustness is defined in terms of numerical stability.

4.2 Effects of Combination Weighting (α)

The main principle of the fusion formula that we introduced is the combination of assessments, which is done on the basis of degrees of certainty:

$$\mu = e + \left[\frac{1}{n}\sum_{i=1}^{n}(\mu_i - e)^{\alpha}\right]^{1/\alpha}, \tag{4.1}$$

where e is the identity of the belief region, and α is the combination weighting of fusion. The former is equal to 0.5 and the latter takes odd values from 1 to ∞. The weighting parameter, α, can be used to emphasize the relative importance of the aggregated bodies of evidence. If α is chosen to be 1, the fusion formula is equivalent to the arithmetic mean, which performs a soft combination with no test for relative certainty. For $\alpha \geq 3$ (α is odd), the combined evidence is pulled toward those pieces of evidence with the smaller degrees of fuzziness. We have thus far chosen $\alpha = 3$. The choice is based on the

Table 4.4: *Results of segmentation for different valuse of* α.

Regions	Ideal	$\alpha = 1$	$\alpha = 3$	$\alpha = 5$	$\alpha = 7$
1	664	586	586	587	587
2	500	474	472	472	472
3	189	118	117	117	117
4	12	3	3	3	5
5	286	206	240	239	239
6	38	40	36	50	52
7	45	4	1	4	1
8	79	86	76	0	0
9	65	62	62	62	0
10	0	4	3	0	3

desire to account for relative certainty in combining evidence, as well as a compromise among supportive and non-supportive pieces of evidence. Now, we focus -o mn.ps muhamad.dvi on the effects of this parameter on the segmentation results. Several values of α are considered. Figure 4.3 shows segmentation results for $\alpha = 1$, $\alpha = 5$, and $\alpha = 7$. Results of image segmentation are compared and evaluated in the next section.

4.2.1 Experimental Evaluation

Segmentation results obtained from different values of combination weighting α are illustrated in Table 4.4. The degrees of similarity between the ideal case and actual cases obtained from the new fusion formula with different values of α are

(1) $s_{if_1} = 90\%$

(2) $s_{if_3} = 91\%$

(3) $s_{if_5} = 88\%$

(4) $s_{if_7} = 88\%$.

The above results show that using $\alpha = 3$ gives the most similar results to the ideal case. In these particular experimentations, it is evident to notice that changing the value of

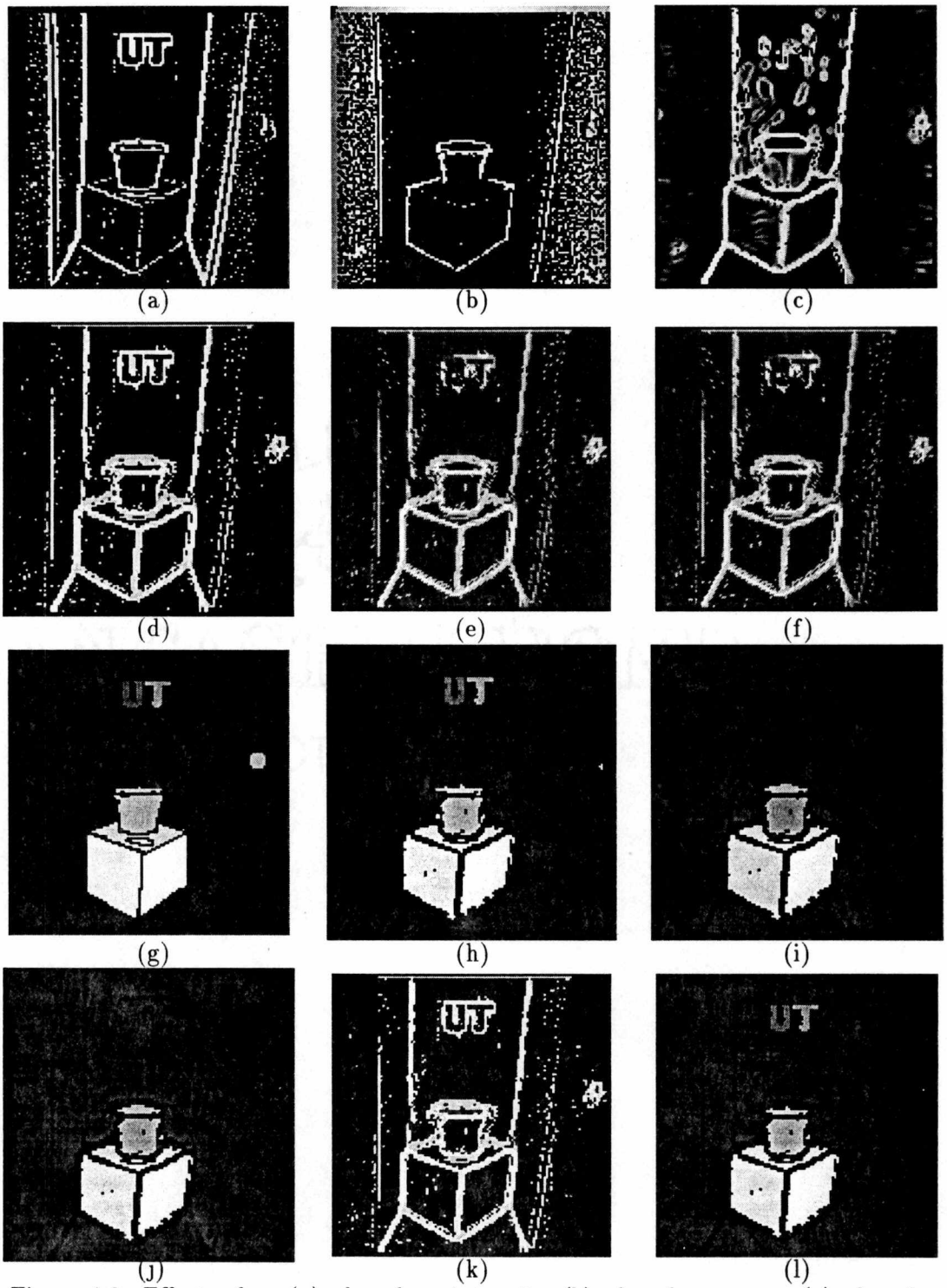


Figure 4.3: *Effects of α: (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d) fused edges ($\alpha = 1$); (e) fused edges ($\alpha = 5$); (f) fused edges ($\alpha = 7$); (g) segmented image of the ideal case (h) segmented image of (d); (i) segmented image of (e); (j) segmented image of (f); (k) fused edges ($\alpha = 3$); (l) segmented image of (k).*

α has no significant effect on the quality of the segmentation results. The reason is that the three edge maps used as inputs to the data fusion system have approximately the same relative certainty. In other words, for any value of α, a compromise among the three sources of information has taken place.

4.3 Effect of the Number of Fuzzy Values (subsets)

Throughout these experimentations, we have quantized universes of discourse into three fuzzy values or subsets. Also, we have used 40% overlapping between fuzzy values. To fuzzify real data, there is no definite procedure by which the selection of the shape of membership functions, the degree of overlapping between the fuzzy subsets, and the number of subsets can be accomplished. In general, the sensitivity of a fuzzy variable determines the number of fuzzy values that are necessary to adequately characterize the entire universe of discourse. Because we have dealt with edge maps of images which show considerable contrast between pixels values of edges and those of background, few fuzzy subsets were needed to characterize the universes of discourse. We have used three fuzzy values. One of these fuzzy values is the concept "moderate," which forms an intermediate region between the fuzzy concepts "weak" and "strong." If this characterization is not flexible enough to encode the knowledge about the system and articulate a complete rule base (RB), then more intermediate regions should be added. In this section, we consider the use of five fuzzy values to represent the universes of discourse. The fuzzy concepts "very weak," "weak," "moderate," "strong," and "very strong" can be represented by membership functions that characterize various levels of edge strength. Figure 4.4 shows membership graphs for the fuzzy subsets of three

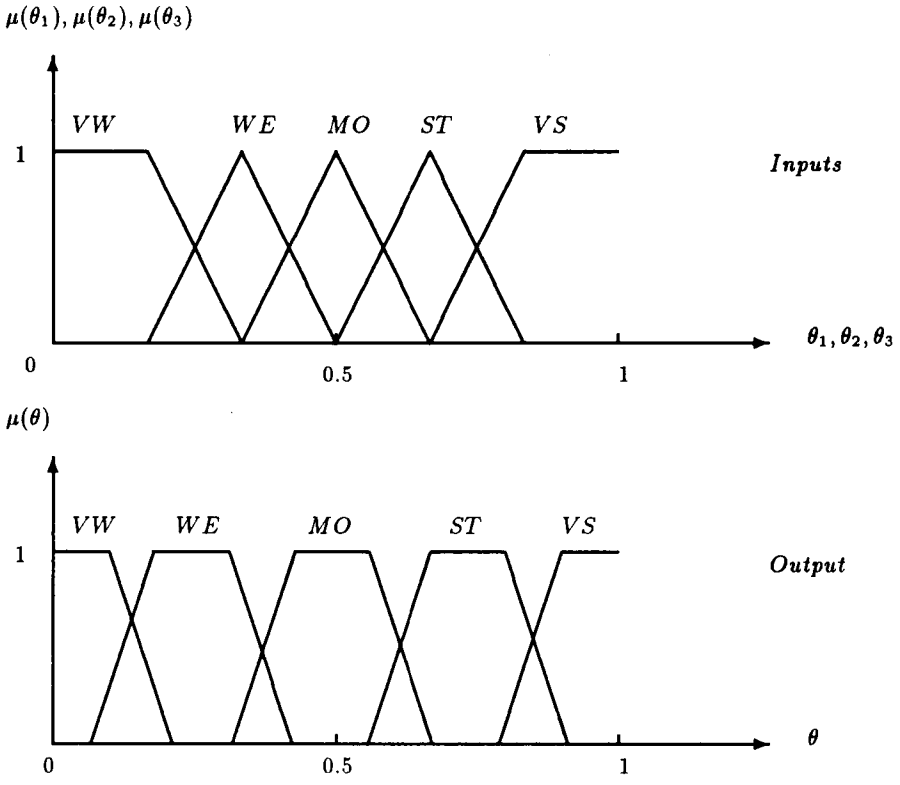


Figure 4.4: *Membership functions for three inputs and one output.*

universes of discourse as inputs and one universe of discourse as an output. Because we have five fuzzy values in each universe of discourse, we can assign a fit vector of length 5, where each element in the fit vector is assigned by the corresponding membership function. This means that for an input value $\theta_i \subset \Theta_i$, we have

$$\theta_i \longrightarrow [\mu_{vw}(\theta_i), \mu_{we}(\theta_i), \mu_{mo}(\theta_i), \mu_{st}(\theta_i), \mu_{vs}(\theta_i)] .$$

Once again, we construct an RB which is able to reflect subjective reasoning when conditional assessments are involved. The constructed RB is able to map input values with different domains to an output value with only one domain. For each location in the scene, there will be three assertions θ_1, θ_2, and θ_3 reported by three sources

of information. In each possible combination, we determine a single output, θ, which suggests an action; *i.e.*,

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(a, b, c) \longrightarrow d\} = \text{elements of } \theta,$$

where a, b, and c are fuzzy values reported by θ_1, θ_2, and θ_3, respectively, and $d \in \theta$ represents a fuzzy output, which is suggested by the combination operator $\oplus$. We assume the following assignments: (1) We index the columns by the five values that partition the edge map from the surface normal of the range image. This input is represented by θ_1. (2) We index the rows by the five values that partition the edge map from the the intensity image. This input is represented by θ_2. To include the edge map from the range image, we fix the fuzzy variable θ_3 as VW, WE, MO, ST, and VS. Entries of the RB are generated from the viewpoint of the practical system operation [75]. Rules are designed by anticipating possible input combinations of θ_1, θ_2, and θ_3 and by applying common sense and our experience in image segmentation to determine the combined edge map θ.

Because roof edges are VS or ST in the normal edge map, the fused edge map must be VS, ST, or at least MO. For instance,

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, WE, WE) \longrightarrow MO\} = \text{ element of } \theta$$

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(ST, WE, ST) \longrightarrow ST\} = \text{ element of } \theta.$$

Because texture edges are VS, ST, or at least MO in the video edge map, and they are VW and/or WE in the normal and range edge maps, the fused edge map must be VS, ST, or at least MO. For instance,

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(VW, MO, VW) \longrightarrow MO\} = \text{ element of } \theta$$

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(WE, ST, WE) \longrightarrow ST\} = \text{ element of } \theta.$$

Because jump edges are VS or ST in the range edge map, the fused edge map must be VS, ST, MO or at least WE. For instance,

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(VW, VW, ST) \longrightarrow WE\} = \text{ element of } \theta$$

$$\theta_1 \oplus \theta_2 \oplus \theta_3 = \{(VW, MO, ST) \longrightarrow MO\} = \text{ element of } \theta.$$

Knowing the problem in hand and using common sense, the rest of the RB is found in a similar fashion. There are 125 possible combinations altogether. Despite the lack of completeness of the RB, the generated rules must be enough to construct a complete inference system that is able to manage uncertainty associated with the fuzzy assessments. The RB of 125 rules is represented by five 5×5 linguistic matrices, which are illustrated in Fig. 4.5. The rest of this procedure is similar to that discussed in the case of three fuzzy inputs. Results from using this fuzzification scheme (*i.e.*, using five fuzzy values) are shown in Fig. 4.6.

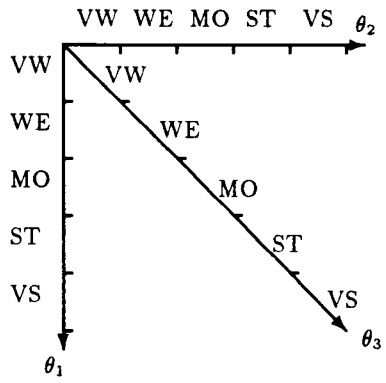
4.3.1 Experimental Evaluation

Segmentation results obtained from the use of five and three fuzzy values are illustrated in Table 4.5. The degrees of similarity between the actual results and ideal results are

(1) $s_{if(3)} = 91\%$

(2) $s_{if(5)} = 91\%$.

The above results show that partitioning the universes of discourse into five fuzzy values has no significant improvement in the final results. These findings were expected because the real inputs (*i.e.*, edge maps) were distinctively clustered such that three fuzzy descriptions were able to characterize three different subsets in their universes of



VW	VW	MO	MO	MO
VW	WE	MO	MO	MO
VW	WE	MO	MO	MO
VW	MO	MO	MO	ST
WE	MO	MO	ST	ST

(a)

WE	MO	MO	ST	ST
WE	MO	MO	ST	ST
MO	MO	MO	ST	ST
MO	MO	ST	ST	ST
MO	ST	ST	ST	VS

(b)

WE	MO	MO	ST	ST
MO	MO	ST	ST	ST
MO	MO	ST	ST	VS
MO	ST	ST	ST	ST
ST	ST	ST	ST	VS

(c)

WE	MO	MO	ST	ST
WE	MO	ST	ST	ST
MO	MO	MO	ST	VS
MO	ST	ST	ST	VS
ST	ST	ST	ST	VS

(d)

WE	MO	ST	ST	VS
MO	MO	MO	ST	ST
MO	MO	ST	ST	VS
MO	ST	ST	VS	VS
ST	ST	VS	VS	VS

(e)

Figure 4.5: *Rule base matrices: columns and rows indexed by the values of θ_1 and θ_2, respectively; (a) $\theta_3 \equiv VW$; (b) $\theta_3 \equiv WE$; (c) $\theta_3 \equiv MO$; (d) $\theta_3 \equiv ST$; (e) $\theta_3 \equiv VS$.*

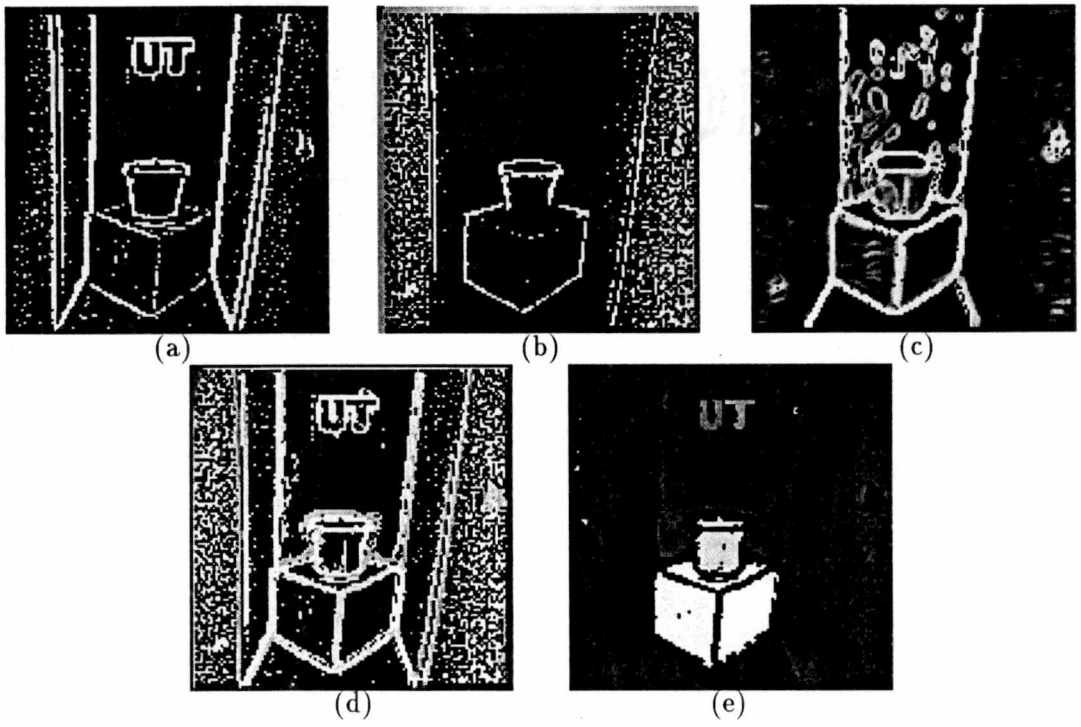


Figure 4.6: *Image segmentation (after data fusion using five fuzzy values): (a) edges from intensity; (b) edges from range; (c) edges from surface normal; (d) fused edges from (a), (b), and (c); (e) segmented image of (d) (10 extracted regions).*

discourse (see Section 2.4.2). We must also point out that the use of five fuzzy values has caused a drastic increase in the size of the RB used in this system and, consequently, the data fusion program requires more time to be executed. In the case of fusing three edge maps, the system requires 2 seconds for execution. In the case of fusing five edge maps, the system requires about 19 seconds for execution. The system used to carry out these implementations is "SPARCstation 2".

Table 4.5: *Results of segmentation (pixels per region).*

Regions	Ideal Case	Fusion: Three Fuzzy Values	Fusion: Five Fuzzy Values
1	664	586	587
2	500	472	474
3	189	117	118
4	12	3	4
5	286	240	241
6	38	36	38
7	45	1	0
8	79	76	76
9	65	62	64
10	0	3	3

CHAPTER 5

Summary and Conclusions

In this dissertation, we have developed a segmentation approach to extract a complete edge map of an arbitrary scene. Real range and intensity images are acquired using the Odetics Laser Range Scanner. Separate sensor modalities do not provide complete information about edges in images. Because combining different sensory modalities helps to capture complete edges in the scene, we have sought the development of a complete data fusion system. Prior to the development of this data fusion system, we have studied a number of methods for combining quantitative assessments provided by various sources of information. We have examined several issues relating to uncertainty management. We have also conducted a comparison between approaches that deal with combining evidence. In view of the difficulties identified in Chapter 2 regarding the fusion of contradictory bodies of evidence, we have pointed out that the way in which pieces of evidence are modeled (fuzzified) plays a crucial role in forming a reliable data fusion system. The Super Bayesian Approach (SBA) aggregates experts' opinions (*likelihoods*) in favor of the decision maker's opinion (*a priori* knowledge). When contradictory assessments between the decision maker and the experts are reported, the SBA allows only for partial compromise among pieces of evidence. The selection of informative or non–informative priors can greatly influence the outcome of combination. In the Bayesian sense, disbelief and the lack of belief are treated similarly.

The mathematical theory of evidence has addressed several problems concerning the

use of Bayes' theory. Distinguishing between disbelief and the lack of belief and combining instead of conditioning contribute to the novelty of this theory. However, consistent overestimation and lack of numerical robustness of the combined evidence often introduce misleading findings. Using the theory of evidence does not incorporate any test for certainty. A complete analysis of several frameworks that deal with uncertainty and combining evidence is presented in Appendices B through D.

We have examined the theory of fuzzy sets and studied its application in data fusion. We have progressed from the concept of modeling real data in terms of fuzzy values to the development of a new fusion formula which satisfies several desirable properties. Some of the desirable properties are critically important in a data fusion system. Significant improvements in the data fusion results achieved using the new data fusion approach have demonstrated the importance of these properties. These findings are based on an experimental evaluation, that compares actual results obtained from this approach, Dempster's rule of combination, and the SBA to the ideal result.

The new fusion formula is different from what has been used traditionally in fuzzy system theories. It tests for certainty and accounts for a relative importance of aggregated bodies of evidence. It is stable, convex, symmetric, monotonic, idempotent, and a quazi–arithmetic mean. We have shown that the lack of some of these desirable properties has degraded the quality of the segmentation results when the SBA or Dempster's rule of combination were used.

Although the development of the fusion formula is the core of this data fusion system, two other components in the system are equally important: (1) modeling data before fusion and (2) recovering data after fusion. The first component represents the process of transforming real or synthetic data, which may be geometrically different,

to a common parametric space from which data can be fused. In this step, we have modeled (fuzzified) different types of data so that a given real measurement is characterized by fuzzy values within a universe of discourse. This fuzzification scheme was able to encode real measurements in terms of fuzzy concepts that are suitable to use in non–deterministic approaches [1]. Fuzzified data of each real input were stored in a data base similar to an ordinary look–up table. Based on this scheme, knowledge about the system is utilized to construct rule bases which are able to articulate fuzzy relations between the inputs and output of the fusion system. The second component represents the process of recovering crisp data (defuzzification) from the combined fuzzy values after fusion. The method of centroid is used to recover crisp data from the cumulative fuzzy outputs. The combined output from each rule represents an entry to the defuzzification step. Although these fuzzification and defuzzification steps have been used recently in "fuzzy control", to the best of our knowledge they are used here for the first time in fuzzy data fusion. To complete the segmentation, the combined edge map was used to isolate homogeneous regions in the scene. Isolated regions are then labeled.

The quality of the segmentation results obtained from the new data fusion approach, Dempster's rule of combination, and the SBA are statistically evaluated against ideal segmentation results for the same scene. The ideal case was established manually. Results from our approach are most similar to those presented in the ideal case. Moreover, we added noise to the input edge maps and repeated the same experimentations. Noise was added using several standard deviation values for the Gaussian filter which was used in the Canny's and Marr's edge detectors (see Section 4.1). Segmentation results from noisy edge maps are also generated and evaluated. Once again, our approach performed better than other approaches.

The effect of parameters, such as the weighting combination and the number of fuzzy values used in each universe of discourse, was studied and evaluated. The increase of the number of fuzzy values has given more flexibility to the fuzzy reasoning process; however, the real time for execution has drastically increased (about 950% of increase).

Fuzzy logic applications have extended to many areas of research, particularly in developing intelligent systems or controls. The former includes areas such as computer vision, pattern recognition, and expert systems. The latter addresses the solution of control system problems. Input values of a control system are used by a fuzzy inference scheme to modulate a control variable that governs the performance of the system. The findings of this work lead us to believe that approaches in the framework of fuzzy set theory are well–suited to solve complex problems such as combining evidence and managing uncertainty. Despite the success of fuzzy logic in many real world applications, few limitations in their use must be pointed out. For a given problem, fuzzification of real data, selection of the shape of membership functions, and the generation of rule bases have no unique solutions. Often, common sense and the experience of fuzzy engineers are important factors in the success of a given design.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. Abdulghafour, J. Goddard, and M. A. Abidi. "Non-deterministic Approaches in Data Fusion - A Review". *SPIE on Sensor Fusion III: 3-D Perception and Recognition*, 1383:596-610, November 1990.

[2] J. Aczel. *Lectures on Functional Equations and Their Applications.* Academic Press, New York, NY 10003, 1966.

[3] J. K. Aggarwal and M. J. Magee. "Determining Motion Parameters Using Intensity Guided Range Sensing". *Pattern Recognition*, 19, No. 2:169-180, April 1986.

[4] P. K. Allen. "A Framework for Implementing Multi-Sensor Robotic Tasks". In *Proceedings of 1987 ASME Conference*, pp. 303-309, New York, NY, 1987.

[5] N. Ayache and O. D. Faugeras. "Building, Registrating, and Fusing Noisy Visual Maps". *The Int'l J. of Robotics Research*, 7, No. 6:45-65, December 1988.

[6] R. Bajcsy and P. Allen. "Converging Disparate Sensory Data". In *Proc. 2nd Int'l Sym. on Robotics Research*, pp. 81-86, Kyoto, Japan, August 1984.

[7] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, NY, 1985.

[8] J. P. Bixler and D. P. Miller. "A Sensory Input System for Autonomous Mobile Robots". In *Spatial Reasoning and Multi-Sensor Fusion: Proceedings of the 1987 Workshop*, pp. 211-219, St. Charles, IL, October 1987.

[9] R. M. Bolle and D. B. Cooper. "On Optimally Combining Pieces of Information, with Application to Estimating 3-D Complex-Object Position From Range Data". *IEEE Trans. on Patt. Anal. Mach. Intell.*, PAMI-8, No. 5:619-638, September 1986.

[10] R. F. Bordley. "A Multiplicative Formula for Aggregating Probability Assessments". *Management Science*, 28, No. 10:1137-1148, October 1982.

[11] J. Canny. "A Computational Approach to Edge Detection". *IEEE Trans. on Patt. Anal. Mach. Intell.*, PAMI-8, No. 6:679-698, November 1986.

[12] D. A. Castanon and D. Teneketzis. "Distributed Estimation Algorithms for Nonlinear Systems". *IEEE J. on Automatic Control*, AC-30, No. 5:418-425, May 1985.

[13] P. S. Chatterjee and T. L. Huntsberger. "Comparison of Techniques for Sensor Fusion Under Uncertain Conditions". In *Proc. of the Sensor Fusion: Spatial Reasoning and Scene Interpretation Conf., Vol. 1003*, pp. 194-199, Cambridge, Ma, 1988.

[14] Y. Cheng and R. L. Kashyap. "A Study of Associative Evidential Reasoning". *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 11, No. 6:623–631, June 1989.

[15] M. D. Cohen, G. Huber, R. L. Keeney, A. H. Levis, L. L. Lopes, A. P. Sage, S. Sen, A. B. Whinston, R. L. Winkler, D. V. Winterfeldt, and L. Zadeh. "Research Needs and the Phenomena of Decisionmaking and Operations". *IEEE Trans. on Sys., Man, and Cyb.*, SMC–15, No. 6:764–775, November/December 1985.

[16] B. deFinetti. *Probability, Induction, and Statistics.* Wiley Publishing, New York, NY, 1972.

[17] B. deFinetti. *Theory of Probability, Vols. 1 and 2.* Wiley Publishing, New York, NY, 1975.

[18] C. J. Delcroix and M. A. Abidi. "Fusion of Range and Intensity Edge Maps". In P. S. Schenker, editor, *Sensor Fusion: Spatial Reasoning and Scene Interpretation, Vol. 1003*, pp. 145–152, Cambridge, MA, November 1988.

[19] A. P. Dempster. "A Generalization of Bayesian Inference". *J. of Royal Statistical Society Series B*, 30:205–247, 1968.

[20] P. Diaconis and S. L. Zabell. "Updating Subjective Probability". *J. of the American Statistical Assoc.*, 77, No. 380:822–830, December 1982.

[21] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications.* Academic Press, New York, NY, 1980.

[22] D. Dubois and H. Prade. *Possibility Theory.* Plenum Press, New York, NY, 1988.

[23] R. O. Duda and D. Nitzan. "Low–Level Processing of Registered Intensity and Range Data". In *The Third International Conference on Pattern Recognition*, pp. 598–601, Coronado, California, November 1976.

[24] R. O. Duda, D. Nitzan, and P. Barrett. "Use of Range and Reflectance Data to Find Planar Surface Regions". *IEEE Trans. on Patt. Anal. Mach. Intell.*, PAMI–1, No. 3:259–271, July 1979.

[25] W. H. DuMouchel and J. E. Harris. "Bayes' Methods for Combining the Results of Cancer Studies in Humans and Other Species". *J. of the American Statistical Assoc.*, 78, No. 382:293–315, June 1983.

[26] H. F. Durrant-Whyte. "Sensor Models and Multisensor Integration". *The Int'l J. of Robotics Research*, 7, No. 6:97–113, 1988.

[27] A. Elfes. "A Tesselated Probabilistic Representation for Spatial Robot Perception and Navigation". In *Presented at the NASA Conf. on Space Telerobotics*, pp. 1–10, Pasadena, CA, February 1989.

[28] A. M. Flynn. "Combining Sonar and Infrared Sensors for Mobile Robot Navigation". *The Int'l J. of Robotics Research*, 7, No. 6:5–14, December 1988.

[29] S. French. "Updating of Belief in the Light of Someone Else's Opinion". *J. R. Statist. Soc. Assoc.*, 143, Part 1:43–48, 1980.

[30] T. D. Garvey. "An Experiment With a System For Locating Objects in Multisensory Images". In *The Third Int'l Joint Conf. on Patt. Recog.*, pp. 567–575, Coronado, California, November 1976.

[31] T. D. Garvey, J. D. Lowrance, and M. A. Fischler. "An Inference Technique for Integrating Knowledge From Disparate Sources". In *Proc. of the 7th Int'l Joint Conf. on Artificial Intelligence*, pp. 319–325, Vancouver, Canada, August 1981.

[32] I. Gath and A. B. Geva. "Unsupervised Optimal Fuzzy Clustering". *IEEE Trans. on PAMI*, July 1989.

[33] S. Geman and D. Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". *IEEE Trans. on Patt. Anal. Mach. Intell.*, PAMI–6, No. 6:721–741, November 1984.

[34] C. Genest. "A Characterization Theorem For Externally Bayesian Groups". *The Annals of Statistics*, 12, No. 3:1100–1105, 1984.

[35] B. Gil, A. Mitiche, and J. K. Aggarwal. "Experiments in Combining Intensity and Range Edge Maps". *Computer Vision, Graphics, and Image Processing*, 21, No. 3:395–411, March 1983.

[36] R. C. Gonzalez and P. Wintz. *Digital Image processing.* Addison–Wesley Publishing Company, Reading, MA, 1987.

[37] P. Grandjean and A. R. D. S. Vincent. "3–D Modeling of Indoor Scenes by Fusion of Noisy Range and Stereo Data". In *The Int'l Conf. on Robotics and Automation*, pp. 681–687, Scottsdale, AZ, May 1989.

[38] S. S. Gupta and J. O. Berger. *Statistical Decision Theory and Related Topics IV.* Springer–Verlag, New York, NY, 1988.

[39] G. Hager and M. Mintz. "Task–Directed Multi–Sensor Fusion". In *The Int'l Conf. on Robotics and Automation*, pp. 662–667, Scottsdale, AZ, May 1989.

[40] P. Hajek. "Combining Functions for Certainty Degrees in Consulting Systems". *Int. J. Man–Machine Studies*, 22:59–76, 1985.

[41] L. O. Hall and A. Kandel. *Designing Fuzzy Expert Systems.* Verlag TUV Rheinland, Germany, 1986.

[42] S. Y. Harmon, G. L. Bianchini, and B. E. Pinz. "Sensor Data Fusion Through a Distributed Blackboard". In *The Int'l Conf. on Robotics and Automation*, pp. 1449–1454, San Francisco, CA, 1986.

[43] H.-Y. Hau and R. L. Kashyap. "On The Robustness of Dempster's Rule of Combination". *IEEE International Workshop on Tools for Artificial Intelligence*, pp. 578–582, October 1989.

[44] R. M. Hogarth. "Cognitive Processes and the Assessment of Subjective Probability Distributions". *J. of the American Statistical Assoc.*, 70, No. 350:271–294, June 1975.

[45] P. J. Huber. *Robust Statistical Procedures.* Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1981.

[46] M. Ishizuka, K. S. Fu, and J. T. P. Yao. "Inference Procedures Under Certainty for the Problem–Reduction Method". *Information Sciences*, 28:179–206, 1982.

[47] R. Jain. "Decisionmaking in the Presence of Fuzzy Variables". *IEEE Trans. on Sys., Man, and Cyb.*, pp. 698–703, October 1976.

[48] H. Jeffreys. *Theory of Probability.* Oxford University Press, London, 1961.

[49] J. M. Jones. *Introduction to Decision Theory.* Richard D. Irwin, Homewood, IL, 1977.

[50] A. Kandel. *Fuzzy Techniques in Pattern Recognition.* John Wiley & Sons, New York, NY, 1982.

[51] G. J. Klir and T. A. Folger. *Fuzzy Sets, Uncertainty, and Information.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[52] B. Kosko. *Neural Networks and Fuzzy Systems.* Prentice Hall, Englewood Cliffs, NJ, 1992.

[53] B. W. Lindgren. *Elements of Decision Theory.* The MacMillan Company, New York, NY, 1971.

[54] D. V. Lindley. "Scoring Rules and the Inevitability of Probability". *Int. Statist., Rev. 50*, pp. 1–26, 1982.

[55] D. V. Lindley. "The Improvement of Probability Judgments". *J. R. Statist. Soc. Assoc.*, 145:117–126, 1982.

[56] D. V. Lindley, a. Tversky, and R. V. Brown. "On the Reconciliation of Probability Assessments". *J. R. Statist. Soc. Assoc.*, 142:146–180, 1979.

[57] J. D. Lowrance, T. D. Garvey, and T. M. Strat. "A Framework for Evidential-Reasoning Systems". In *Proc. Nat'l Conf. on Artificial Intelligence*, pp. 896–903, Los Altos, CA, 1986.

[58] R. C. Luo, M.-H. Lin, and R. S. Scherp. "The Issues and Approaches of a Robot Multi-Sensor Integration". In *Proc. IEEE Int'l Conf. on Robotics and Automation*, pp. 1941–1946, Raleigh, NC, March 1987.

[59] M. J. Magee, B. A. Boyter, C.-H. Chien, and J. K. Aggarwal. "Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects". *IEEE Trans. on Patt. Anal. Mach. Intell.*, PAMI-7, No. 6:629–637, November 1985.

[60] R. C. Mann. "Multi-Sensor Integration Using Concurrent Computing". In *SPIE Symposium on Infrared Sensors and Sensor Fusion*, Orlando, FL, May 1987.

[61] D. C. Marr and E. Hildreth. "Theory of Edge Detection". *Proc. Roy. Soc., London*, B-207:187–217, 1980.

[62] K. J. McConway. "Marginalization and Linear Opinion Pools". *J. of the American Statistical Assoc.*, 76, No. 374:410–414, June 1981.

[63] S. K. Pal. "Fuzzy Tools for the Management of Uncertainty in Pattern Recognition, Image Analysis, Vision, and Expert Systems". *International Journal of Systems Science*, pp. 511–549, 1991.

[64] H. Prade. "A Computational Approach to Approximate and Plausible Reasoning with Applications to Expert Systems". *IEEE Trans. on Patt. Anal. Mach. Intell.*, PAMI-7, No. 3:260–283, May 1985.

[65] H. Raiffa. *Decision Analysis: Introductory Lectures on Choices Under Certainty.* Addison-Wesley, Reading, MA, 1968.

[66] S. A. Schmitt. *Measuring Uncertainty.* Addison-Wesley, Reading, MA, 1969.

[67] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ, 1976.

[68] G. Shafer. "Belief Functions and Parametric Models". *J. R. Statist. Soc. B*, 44, No. 3:322–352, 1982.

[69] G. Shafer. "Lindley's Paradox". *J. of the American Statistical Assoc.*, 77, No. 378:325–351, June 1982.

[70] S. Shekhar, O. Khatib, and M. Shimojo. "Object Localization with Multiple Sensors". *The Int'l J. of Robotics Research*, 7, No. 6:34–44, December 1988.

[71] D. Sher. "Optimal Likelihood Generators for Edge Detection Under Gaussian Additive Noise". In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 94–99, Miami Beach, FL, June 1986.

[72] D. Sher. "Evidence Combination Using Likelihood Generators". In *Proc. DARPA Image Understanding Workshop*, pp. 655–662, Los Angeles, CA, February 1987.

[73] W. Silvert. "Symmetric summation: A Class of Operations on Fuzzy Sets". *IEEE Trans. Syst., Man. Cybern.*, 9:657–659, 1979.

[74] J. C. Sluder and M. A. Abidi. "Parallel Range Data Processing: A Real Case Study". *SPIE Intelligent Robots and Computer Vision X*, 1608:248–261, November 1991.

[75] G. C. D. Sousa and B. K. Bose. "A Fuzzy Set Theory Based Control of A Phased–controlled Converter DC Machine Drive". *IEEE/IAS Annual Meeting*, October 1991.

[76] M. Sugeno. *"Theory of Fuzzy Integrals and Its Applications"*. PhD thesis, Tokyo Institute of Technology, Japan, 1974.

[77] C. G. Wagner. "Consensus for Belief Functions and Related Uncertainty Measures". *Theory and Decision*, 26:295–304, 1989.

[78] P. P. Wang and S. K. Chang. *Fuzzy Sets.* Plenum Press, New York, NY, 1980.

[79] Y. F. Wang and J. K. Aggarwal. "On Modelling 3–D Objects Using Mutliple Sensory Data". In *1987 Robotics and Automation Conference*, pp. 1098–1103, Raleigh, NC, March 1987.

[80] T. Whalen. "Decisionmaking Under Uncertainty with Various Assumptions about Available Information". *IEEE Trans. on Sys., Man, and Cyb.*, SMC–14, No. 6:888–900, November/December 1984.

[81] G. Wolberg and T. Pavlidis. "Restoration of Binary Images Using Stochastic Relaxation with Annealing". *Pattern Recognition Letters 3*, pp. 375–388, December 1985.

[82] R. R. Yager. "Approximate Reasoning as a Basis for Rule–Based Expert Systems". *IEEE Trans. on Sys., Man, and Cyb.*, SMC–14, No. 4:636–643, July/August 1984.

[83] R. R. Yager. "Q–Projections on Possibility Distributions". *IEEE Trans. on Sys., Man, and Cyb.*, SMC–15, No. 6:775–777, November/December 1985.

[84] L. A. Zadeh. "Fuzzy Sets". *Information and Control*, 8:338–353, 1965.

[85] L. A. Zadeh. "Probability Measures of Fuzzy Events". *J. Math. Analysis and Appl.*, 10:421–427, 1968.

[86] L. A. Zadeh. "Similarity Relations and Fuzzy Orderings". *Information Sciences*, 3:177–200, 1971.

[87] L. A. Zadeh. "A Fuzzy–Set–Theoretic Interpretation of Linguistic Hedges". *J. of Cyb.*, 2, No. 2:4–34, 1972.

[88] L. A. Zadeh. "The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Parts 1 and 2". *Information Sciences*, 8:199–249, 301–357, 1975.

[89] L. A. Zadeh. "A Fuzzy–Algorithmic Approach to the Definition of Complex or Imprecise Concepts". *Int'l J. Man–Machine Studies*, 8:249–291, 1976.

[90] L. A. Zadeh. "A Computational Approach to Fuzzy Quantifiers in Natural Language". *Comp. and Maths with Appls.*, 9:149–184, 1983.

[91] L. A. Zadeh. "The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems". *Fuzzy Sets and Systems*, 11:199–227, 1983.

[92] L. A. Zadeh. *Fuzzy Sets and Applications: Selected Papers by L. A. Zadeh.* John Wiley & Sons, New York, NY, 1987.

[93] H.-J. Zimmermann. *Fuzzy Sets, Decision Making, and Expert Systems.* Kluwer Academic Publishers, Boston, MA, 1987.

APPENDICES

APPENDIX A

Modeling Uncertainty: The Probability Theory

A.1 Probabilities: Concepts and Properties

Loosely speaking, a probability is a number that represents the relative chance that some event will occur. This relative chance is given by a number lying between 0 and 1. A probability of 0 indicates that an event will never occur or is impossible. A probability of 1 indicates that an event is certain to occur. Probabilities of values lying between 0 and 1 indicate that events are uncertain to occur. These types of probabilities have had a great deal of attention in decision theory.

Formally speaking, the probability (p) that an event θ will occur is denoted by $p(\theta)$. The general property of probability is that probability for any event, θ, lying between 0 and 1 is expressed as

$$0 \leq p(\theta) \leq 1 . \tag{A.1}$$

By definition, the probability of a certain event, θ, is 1:

$$p(\theta) = 1 . \tag{A.2}$$

Another important property is that if θ_1 and θ_2 are two mutually exclusive events, then

$$p(\theta_1 \text{ or } \theta_2) \quad = \quad p(\theta_1) + p(\theta_2) . \tag{A.3}$$

This is a useful property which allows us to compute the probability of composite events if the probabilities of elementary events are known. An obvious extension holds for any

number of mutually exclusive events [49]. Finally, if two events θ and $\overline{\theta}$ are defined as complements, then the sum of the probabilities $p(\theta)$ and its negation $p(\overline{\theta})$ is 1; that is,

$$p(\theta) + p(\overline{\theta}) \;=\; 1\,, \tag{A.4}$$

where $p(\theta)$ and $p(\overline{\theta})$ are assumed to be mutually exclusive and collectively exhaustive. Mutually exclusive means that if one event occurs, then the other will not occur. Collectively exhaustive means that one event has to occur [17].

A.2 Probability Rules for Generating Compound Events

In the previous section, we discussed some ideal properties by which one can compute probabilities of events that may occur. In real world situations such computations are not so easy. We present some rules by which probabilities of more complex situations can be derived. These rules are of great importance, particularly, when other types of probability distributions are introduced in subsequent sections.

First, assume a conceptual chance experiment. Then, we define a sample space to be the set of all probable outcomes of this experiment. We denote the sample space as Θ and its set of elements as $\{\theta_1, \theta_2, \ldots, \theta_n\}$. The individual elements are often called events.

There are two basic rules by which two or more mutually exclusive and collectively exhaustive events can be combined. First, the "or" rule is

$$p(\theta_1 \text{ or } \theta_2|\Theta) \;=\; p(\theta_1|\Theta) + p(\theta_2|\Theta)\,, \tag{A.5}$$

where $p(\theta_1|\Theta)$ and $p(\theta_2|\Theta)$ are the *a posteriori* probabilities for θ_1 and θ_2 given Θ, respectively. This means that the probability of one or the other of the two happening

is the sum of their separate probabilities. In this rule, which one happens first is of no importance. From the definition of exhaustiveness, one event has to occur, and we can write the following result:

$$p(\theta_1 \text{ or } \theta_2 \text{ or } \ldots \theta_n|\Theta) = p(\theta_1|\Theta) + p(\theta_2|\Theta) + \ldots + p(\theta_n|\Theta) = 1\,. \qquad \text{(A.6)}$$

Again, this is the basic property that the sum of probabilities in a probability distribution is 1 (sometimes we write Eq. A.6 as $p(\Theta) = 1$, where Θ is the sample space that contains all probable outcomes). Secondly, the "and" rule is

$$p(\theta_1 \text{ and } \theta_2|\Theta) = p(\theta_1|\Theta) + p(\theta_2|\Theta) - p(\theta_1 \text{ or } \theta_2|\Theta)\,. \qquad \text{(A.7)}$$

This equation states that the probability of both events, θ_1 and θ_2, occurring at the same time is the probability of the intersection of these two events. This intersection contains points that are common to both θ_1 and θ_2. An obvious observation from Eq. A.7 is that $p(\theta_1 \text{ and } \theta_2|\Theta) = 0$ if θ_1 and θ_2 are assumed to be mutually exclusive (disjoint) events.

A.3 Joint and Conditional Probability

In many situations, we may not be able to represent probabilities of events in terms of their probable elementary occurrences. Fortunately, knowing partial information committed to some events allows us to deduce some conditional probabilities committed to other events. It is important to emphasize that conditional probability is a defined concept which can be expressed in terms of joint probabilities. To do so, let A_1, A_2, ..., A_m and, similarly, B_1, B_2, ..., B_n denote disjoint events associated with an experiment $\mathcal{B}$. Then, the probability of some event, say A_i, given that some other event, say B_j, has occurred, is called the conditional probability and denoted by $P(A_i|B_j)$. Note that

the probability array $P(A_i \cap B_j)$, which contains all probable combinations for every $i = 1, \ldots, m$ and $j = 1, \ldots, m$, is called the joint probability array. The conditional probability of A_i given B_j is defined as

$$P(A_i|B_j) = \frac{P(A_i \cap B_j)}{P(B_j)}. \tag{A.8}$$

Similarly, the conditional probability of B_j given A_i is defined as

$$P(B_j|A_i) = \frac{P(B_j \cap A_i)}{P(A_i)}. \tag{A.9}$$

It is important to notice that the quantities $P(B_j)$ and $P(A_i)$ are introduced in Eqs. A.8 and A.9 for normalization reasons; they satisfy the basic property of probability distribution. The sum of the new sets after normalization will add up to unity. The functions $P(A_i)$ and $P(B_j)$ are usually called the marginal probabilities. The usefulness of the joint and conditional density functions of A_i and B_j is that one can retrieve the individual density functions of A_i and B_j.

A.4 Independence

Two events are said to be independent if the occurrence of one event does not affect the occurrence of the second. Formally, if events A and B are independent, then

$$P(A \cap B) = P(A)P(B). \tag{A.10}$$

Also, as a result of independency between events A and B, the conditional probabilities in Eqs. A.8 and A.9 become

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B). \end{aligned} \tag{A.11}$$

A.5 Likelihood Functions

A likelihood function is also a defined concept as in the case of conditional probability. It is exactly a conditional probability given a particular alternative. For observed data, x, the function $p(x|\theta)$ is called the likelihood function, where $p(x|\theta)$ is considered a function of θ. In this defined concept, only the actual observed data is relevant to evidence about θ. If $p(x|\theta)$ were large, the observed value, x, of an experiment is more probable to occur, and θ is more likely to be true. In contrast, if $p(x|\theta)$ were small, the observed value, x, of an experiment is less probable to occur and θ is less likely to be true [7].

A.6 Bayes' Theorem

Bayes' theorem is a tool by which one can compute the probability of the occurrence of an event in the future by observing the probability of the occurrence of similar events in the past. More specifically, a probability distribution that describes the occurrence of an event in the past is defined as the *a priori* probability. Then we run an experiment and record the observations that are used to modify the prior probability distribution and obtain a new one. This is called the *a posteriori* distribution. Let us assume that there is a sample space Θ, which contains only two populations, $\Theta = \{\theta_1, \theta_2\}$, from which the sample x is observed. The problem is to compute the probability that the sample x was drawn from θ_1. From the probability theory, if two events intersect with each other such as x and θ_1, one may apply Eq. A.8 as follows:

$$p(\theta_1|x) = \frac{p(\theta_1 \cap x)}{p(x)}, \tag{A.12}$$

where

$$p(x) = p(\theta_1 \cap x) + p(\theta_2 \cap x). \tag{A.13}$$

From Eq. A.12 we can readily write

$$p(x|\theta_1) = \frac{p(x \cap \theta_1)}{p(\theta_1)}, \tag{A.14}$$

or

$$p(x \cap \theta_1) = p(x|\theta_1) \cdot p(\theta_1). \tag{A.15}$$

Similarly,

$$p(\theta_2 \cap x) = p(x|\theta_2) \cdot p(\theta_2). \tag{A.16}$$

Substituting Eqs. A.15 and A.16 into Eq. A.13, we get

$$p(x) = p(x|\theta_1) \cdot p(\theta_1) + p(x|\theta_2) \cdot p(\theta_2). \tag{A.17}$$

Substituting Eqs. A.17 and A.15 into Eq. A.12, we finally get

$$p(\theta_1|x) = \frac{p(x|\theta_1) \cdot p(\theta_1)}{p(x|\theta_1) \cdot p(\theta_1) + p(x|\theta_2) \cdot p(\theta_2)}, \tag{A.18}$$

where $p(\theta_1|x)$ is called the probability that the sample x comes from the population θ_1, which is often referred to as the *a posteriori* probability of the sample x. The functions $p(\theta_1)$ and $p(\theta_2)$ are the prior probability distributions of the populations θ_1 and θ_2, respectively. The probabilities $p(x|\theta_1)$ and $p(x|\theta_2)$ are the likelihood functions of the populations θ_1 and θ_2, respectively. The likelihood functions are the carriers of information by which the prior probability distribution can be modified. Of course, larger $p(x|\theta_1)$ and $p(x|\theta_2)$ increase the degree of possible occurrence of the observed

variables. The case of two populations can be readily extended to the case of a sample space Θ, which contains n populations: $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. The posterior probability for n populations is given by

$$p(\theta_i|x) = p(x|\theta_i)p(\theta_i)/\sum_{j=1}^{n} p(x|\theta_j) \cdot p(\theta_j), \qquad \text{(A.19)}$$

where $i = 1, 2, \ldots, n$, so that for every subscript of θ, we would have a related *a posteriori* probability. Note that the denominator of Eq. A.19 is nothing more than the total probability. Thus, using Eq. A.19, one can compute the revised assessment concerning the occurrence of x.

A.7 Subjective and Prior Probabilities

In many decision problems, and especially for those that use Bayesian techniques, the prior probabilities are very important quantities, but are difficult to determine. Before discussing different types of priors, it is appropriate to present some other probabilistic elements that have similar probabilistic concepts. In the next subsections, methods of constructing priors will be discussed in more detail.

I. Subjective Probabilities

The concept of subjective probability is based on the judgmental inference which reflects some personal guesses regarding the occurrence of events even though frequency probability cannot be assigned to the events in question. An illustrative example for the subjective probability is if we deduce the probability of a certain event $p(\theta)$ by comparing it with the probability of its negation $p(\overline{\theta})$. If, for example, the event θ is three times as likely to occur as $\overline{\theta}$, then obviously $p(\theta) = 3/4$ and $p(\overline{\theta}) = 1/4$. Another way of deducing the subjective probability is to use utility functions [7, 16, 54]. A zero

utility function allows us to deduce a "fair" subjective probability, $p(\theta)$, for a certain event θ. In a gambling game, for instance, let n be lost if θ occurs, and $1-n$ be gained if $\overline{\theta}$ occurs. Assuming a fair game (*i.e.*, the expected utility of the gamble is zero), the utility function becomes

$$0 = u(-n)p(\theta) + u(1-n)(1-p(\theta))\,.$$

Solving for $p(\theta)$, we get

$$p(\theta) \;=\; \frac{u(1-n)}{u(1-n)-u(-n)}\,. \tag{A.20}$$

If n is "small," u is approximately linear; then Eq. A.20 becomes

$$p(\theta) \;\cong\; (1-n)\,. \tag{A.21}$$

II. Prior Probability Through Subjective Determination

Let Θ be a sample space containing all possible events and $\pi(\theta)$ be the prior density. If Θ is discrete, then the prior density can be determined easily by one of the methods of determining the subjective probability. If Θ is continuous, then the problem of computing $\pi(\theta)$ becomes more difficult. Several techniques may be used for this purpose, such as the histogram approach and the relative likelihood approach. There are other approaches to solve this problem; however, only two approaches will be discussed. Other methods may be found in [7], from which other useful references can be obtained.

(a) The Histogram Approach: In this approach, the continuous, Θ, is divided into intervals, the subjective probability is obtained for each interval, then the probability histogram is plotted. From the plotted histogram, a smooth probability density function $\pi(\theta)$ can be sketched. There are no rules for how many intervals are used in the histogram approach. However, more intervals are needed if more detailed information

about the problem in question is required. In this case, the amount of computation increases and a balance of this trade–off must be considered.

(b) The Relative Likelihood Approach: If Θ is continuous, then one may consider the likelihood information to deduce a subjective density which can be considered as a prior density. This is an intuitive approach, which is accomplished by picking various points in the given likelihoods and directly sketching an approximate prior. One must check, however, for consistency while sketching and normalizing to meet the basic properties of probability.

A.8 Noninformative Priors

From a statistical point of view, the availability of some knowledge about a given event has a tremendous impact on the credibility of the obtained results. Bayes' rules are very effective when reliable prior knowledge is available. Unfortunately, in real world problems, we do not have these luxuries. We are forced to make some subjective decisions regarding prior information. Due to this weakness of prior assumptions, Bayesians are heavily criticized [7]. Nevertheless, Bayesian mechanisms are still attractive to many statisticians. Here, the idea of noninformative priors arises as a remedy to the problem of Bayes' rules. Several cases can be stated in this regard. For instance, Bayesians represent *ignorance* by $\pi(\theta) = 1/2$, which gives no favor to either of two hypotheses concerning θ (*i.e.*, the likelihood is the only proposition to be considered) [7], and certainly $\pi(\theta)$ is a noninformative prior.

Another illustrative example can also show a noninformative prior. If we let the parameter in question θ be normal, then the parameter space is $\Theta = [-\infty, \infty]$. This

would suggest a noninformative prior that gives equal weight to all possible values of θ, that is, $\pi(\theta) = K > 0$. Unfortunately, this would violate the basic property of probability density, which leads to an infinite mass [7], *i.e.*,

$$\int \pi(\theta) d\theta = \infty.$$

Nevertheless, this prior can be utilized by simply choosing $\pi(\theta) = 1$, where the choice of K is unimportant. This prior is usually called the uniform prior density, which was introduced by Laplace (1812). It must be used cautiously because constant priors show inconsistency under transformation. In other words, they suffer from a lack of invariance under transformation.

General settings for noninformative priors were introduced by Jeffrey [48] to choose a noninformative prior to be one of the following forms

$$\pi(\theta) = [I(\theta)]^{1/2}, \tag{A.22}$$

where $I(\theta)$ is the expected Fisher information given by

$$I(\theta) = -E_\theta \left[\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2} \right].$$

If $\Theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$ is a vector in R^p, Jeffreys [48] suggested that

$$\pi(\theta) = [det \mathbf{I}(\Theta)]^{1/2}, \tag{A.23}$$

where $\mathbf{I}(\Theta)$ is the $(P \times P)$ Fisher information matrix. For a single element, Fisher information is given by

$$\mathbf{I}_{i,j}(\Theta) = -E_\Theta [\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(X|\Theta)].$$

To see how this method works, let us consider the following example. Suppose the location-scale density is of the form $\delta^{-1} f((x-\theta)/\delta)$, where $\theta \in R^1$ and $\delta > 0$ are the

unknown parameters. For simplicity, suppose that the location–scale density is $\mathcal{N}(\theta, \delta^2)$. In this case $\Theta = (\theta, \delta)$, and the Fisher information matrix is

$$\begin{aligned}
\mathbf{I}(\Theta) &= -E_\Theta \begin{bmatrix} \frac{\partial^2}{\partial\theta^2}(-\frac{(X-\theta)^2}{2\delta^2}) & \frac{\partial^2}{\partial\theta\partial\delta}(-\frac{(X-\theta)^2}{2\delta^2}) \\ \frac{\partial^2}{\partial\theta\partial\delta}(-\frac{(X-\theta)^2}{2\delta^2}) & \frac{\partial^2}{\partial\delta^2}(-\frac{(X-\theta)^2}{2\delta^2}) \end{bmatrix} \\
&= -E_\Theta \begin{bmatrix} -1/\delta^2 & 2(\theta - X)/\delta^3 \\ 2(\theta - X)/\delta^3 & -3(X-\theta)^2/\delta^4 \end{bmatrix} \\
&= \begin{bmatrix} 1/\delta^2 & 0 \\ 0 & 3/\delta^2 \end{bmatrix}.
\end{aligned}$$

Equation A.23 becomes

$$\pi(\Theta) = (\frac{1}{\delta^2} \cdot \frac{3}{\delta^2})^{1/2} \propto \frac{1}{\delta^2}.$$

This is also an improper density, which violates the basic property of probability. As an adjustment to this result, this density could be derived by assuming independence of θ and δ and multiplying the result obtained above by δ; *i.e.*,

$$\pi(\theta, \delta) = \frac{1}{\delta}, \tag{A.24}$$

which is the noninformative prior for a location–scale problem that is recommended by Jeffreys [48].

To many probabilists, Bayes' rules still, in many cases, have serious problems. They argue that noninformative priors are often unacceptable for several reasons. One is that most methods of deriving noninformative priors rely on some experimental structure that may violate the principle of likelihood. A second reason is that noninformative priors perform poorly in many situations [7]. From a statistical point of view, it is rare to have a technique in statistical decision theory that could perform perfectly in any given case. Therefore, this problematic matter encourages many researchers to explore ways for better performances.

APPENDIX B

Modeling Uncertainty: The Evidence Theory

The theory of evidence deals with the weight of evidence and with an assigned degree of support based on evidence. Like most theories and techniques dealing with uncertainty, the assigned degree of support is between zero and one to indicate the strength of support committed to a proposition. Unlike other theories, this attempt (*i.e.*, the theory of evidence) is relatively more general than all previous attempts in statistical decision theory [67]. The generality emerged from Dempster's rule of combination, which pools several bodies of evidence committed independently to a single event. In other words, it is more realistic in dealing with uncertainty because neither likelihoods nor informative or noninformative priors are needed. It simply treats pieces of evidence according to the level of confidence and degree of support in each one. This theory was established on the fact that each source of information delivers only a partial belief about a proposition. The total belief is best deduced by using Dempster's rule of combination, which fairly pools different sources of information. Notions presented in this section are taken from [22, 31, 57, 67, 68, 69, 77].

B.1 Propositions and the Frame of Discernment

Before introducing the belief functions, it is necessary to define some notations that will be used. The notation is similar to that used in the discussion of the probability theory. Let Θ be the frame of discernment (*i.e.*, sample space) and $\{\theta_1, \theta_2, \ldots, \theta_n\}$ be

the set of all its subsets. The event θ is often called a proposition. If θ is allowed to be any arbitrary parameter, possibly non-numerical, then θ and Θ can be chosen so that the set 2^{Θ} includes all the particular fixed sets of propositions. For instance, if $\Theta = \{\theta_1, \theta_2\}$ is the set of possibilities, then $2^{\Theta} = \{\emptyset, \Theta, \{\theta_1\}, \{\theta_2\}\}$ is the set of its subsets. Note that the empty set, $\emptyset$, corresponds to a proposition which is known to be false, and Θ corresponds to a proposition that is known to be true. In other examples, 2^{Θ} may include non-numerical values such as date and/or place of origin assigned for θ, and Θ can be the set of all pairs consisting of a possible date and a possible place of origin. Therefore, the frame of discernment can be conceptually thought of as what we know or think we know about the problem in question, and it is assumed to be finite.

B.2 Belief Functions

A portion of belief committed to a proposition is also committed to any other proposition that it may imply. Suppose A is a subset of a frame Θ. Then, a portion of the belief committed to A is also committed to any subset containing A. Thus, some of the total belief committed to A may also be committed to any subset contained in A and the rest is exactly committed to A. Furthermore, if a portion of belief is committed to a proposition A, the rest of this belief need not be committed to its negation $\overline{A}$. The latter is a fundamental deviation from the probability theory. Therefore, partitioning the total belief to portions among subsets of a frame of discernment is possible and can be represented by basic probability numbers.

Definition: If Θ is a frame of discernment and $A \subset \Theta$, then a function $m : 2^{\Theta} \rightarrow [0, 1]$ is called a basic probability assignment whenever

1. $m(\emptyset) = 0$,

2. $\sum_{A \subset \Theta} m(A) = 1$.

Condition (1) says that no belief should be committed to the empty subset $\emptyset$, while condition (2) says that the total belief must add up to one. Note that $m(A)$ is the belief that one exactly commits to A, and not the total belief committed to A. To obtain the total belief committed to A, it is necessary to include a new subset, say $B \subset A$. Therefore, the total belief committed to A is

$$Bel\,(A) \;=\; \sum_{B \subset A} m(B)\,. \tag{B.1}$$

A function given by Eq. B.1 is called a belief function over Θ for some basic probability assignment $m : 2^{\Theta} \rightarrow [0,1]$. In this case, if $m(A) > 0$, then A is called a focal element of this belief function.

Theorem 2.1 [67]: Suppose Θ is a frame of discernment. Then a function Bel: $2^{\Theta} \rightarrow [0,1]$ is a belief function if and only if it 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,

$$Bel(A_1 \cup \ldots \cup A_n) \geq \sum_{I \subset \{1,\ldots,n\}, I \neq \emptyset} (-1)^{|I|+1} \cdot Bel(\bigcap_{i \in I} A_i)\,.$$

Condition (3) says that the total belief committed to the collection $A_1, \ldots, A_n$ of subsets of Θ is equal to or less than the total belief committed to the union of $A_1, \ldots, A_n$. The

symbol $|I|$ denotes the cardinality of the set I, *i.e.*, the number of elements in I.

Theorem 2.2 [67]: Suppose Bel: $2^{\Theta} \rightarrow [0,1]$ is the belief function given by the basic probability assignment $m : 2^{\Theta} \rightarrow [0,1]$. Then

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} Bel(B). \tag{B.2}$$

This theorem guarantees that the basic probability assignment for a given belief function is unique and can be computed from the belief function.

B.3 Lower and Upper Probabilities: Credibility and Plausibility

Suppose $A \subset \Theta$ and the total belief committed to A is $Bel(A)$. Clearly, $Bel(A)$ does not show the degree of doubt committed to A. That is, the belief committed to its negation is not known yet. A relation which can fully describe both the degree of belief and the degree of doubt is given by the following equation.

$$Dou(A) = Bel(\overline{A}).$$

A more general description is the upper probability or often called the plausibility, and it is given by

$$P\ell(A) = 1 - Bel(\overline{A}).$$

The plausibility measure $P\ell(A)$ establishes the maximum likelihood that could be committed to a proposition A. The lower probability, $Cr(A)$, called credibility, represents the minimum likelihood that could be committed to a proposition, A. The latter is often called the degree of support for the proposition A and is given by

$$Cr(A) = Bel\ (A) = \sum_{B \subset A} m(B). \tag{B.3}$$

Furthermore, the degree of uncertainty about the actual probability value for a proposition is determined by the width of its evidential interval; *i.e.*,

$$U(A) \;=\; P\ell(A) - Cr(A)\,, \tag{B.4}$$

where $U(A)$ is the uncertainty of A. If $U(A)$ is equal to 0, the system is Bayesian [22, 31, 57, 67]. A common use of the "evidential propositional calculus" is the notation $A_{[Cr(A),P\ell(A)]}$, which represents the evidential interval of A. Several references related to the theory of evidence (belief) can be consulted [19, 67, 68, 69].

B.4 Bayesian as a Special Case of the Theory of Evidence

As given in Eq. B.4, if the uncertainty of A is zero, then the system is Bayesian. Formally, this can be defined as follows:

Definition: If Θ is a frame of discernment, then a function, Bel: $2^{\Theta} \rightarrow [0,1]$, is called a Bayesian belief function if

1. $Bel(\emptyset) = 0$,
2. $Bel(\Theta) = 1$,
3. $Bel(A \cup B) = Bel(A) + Bel(B)$ whenever $A, B \subset \Theta$ and $A \cap B = \emptyset$ (*i.e.*, mutually exclusive.)

This definition implies that $U(A) = 0$; that is, $Cr(A) = P\ell(A)$ and $Bel(A) + Bel(\overline{A}) = 1$ for all $A \subset \Theta$ [67, 22].

APPENDIX C

Modeling Uncertainty: The Possibility Theory

The possibility theory was introduced by L. Zadeh [86]. This concept of fuzzy sets is motivated by the need to establish systematic methods for dealing with lexical elasticity and imprecise or uncertain quantities, which are not probabilistic in nature. The word "possibility," however, implies some flexible degree of support viewed not as an alternative to probability, but as a method that is able to address issues without which they are not solvable [86]. It offers a representation for quantitative judgment, which is considered a limiting case of the qualitative one. The possibility theory is able, at the same time, to represent imprecision in terms of fuzzy sets and to quantify uncertainty through two proposed measurements, the possibility and necessity [22]. In the following, we will discuss in more detail the concepts of possibility theory including measures of possibility and the relation between the possibility theory and other theories, all of which are in the framework of the theory of fuzzy sets.

C.1 The Fuzzy Sets Theory

This presentation of fuzzy sets introduces basic definitions and properties related to the theory of fuzzy sets as they will be used later to model uncertainty and imprecision. An introduction is given to show reasoning in uncertain (*i.e., fuzzy*) environments. The theory of fuzzy sets, introduced by Zadeh in 1965 [84], deals primarily with inexact, incomplete, or uncertain quantities. Dealing with real world problems often presents

situations that crisp mathematics, with its usual ideal assumptions, is unable to handle. The argument is initiated by the fact that most statistical problems are ill–defined, so that a fuzzified version of any statistical concept is a more effective tool in decision–making processes [15]. Even though the fuzzy sets theory is basically similar to human intuition in dealing with uncertainty, fuzzy mathematics allows for a better mechanism by which a systematic inference procedure can take place [47]. The concept of membership grade was introduced to represent fuzzy confidence or the fuzzy degree of support for a particular member (*i.e.*, proposition) in the universe of discourse (*i.e.*, sample space). In the following subsections, definitions of the basic notions within the language of fuzzy sets and the application validity of such a conceptual framework will be discussed.

The theory of fuzzy sets is a generalization of abstract set theory. It was shown that the theory of fuzzy sets can deal with even crisp sets. As a consequence, the theory of fuzzy sets has become more and more attractive. As mentioned earlier, this technique of representing uncertainty is similar to human intuition, which can give wider flexibility and requires no fixed absolute constraints. For instance, we define a set, Θ, as the set of men. Formally,

$$\Theta = \{\theta | \theta \text{ is a man } \}.$$

We assign a degree of uncertainty describing these men, say tall men. The word, "tall," is clearly a fuzzy description because a tall man might be considered above 6, 6.5, or 7 feet. These fuzzy descriptions of three types of tall men can be represented by utilizing the concept of the membership grade. The variety of representations for the same word, "tall," using the membership grade makes this theory more effective than most other theories when analyzing and reasoning about uncertain or imprecise quantities.

Definition: If $\Theta = \{\theta\}$ denotes a space of objects, then a fuzzy set, A, in Θ is a set of ordered pairs,

$$A = \{(\theta, \mu_A(\theta))\} \ , \theta \in \Theta \, , \tag{C.1}$$

where μ_A is the membership function that maps Θ to the membership space M, and $\mu_A(\theta)$ is the grade of membership of θ in A. Note that if M contains only the two points 0 and 1, then A is crisp and μ_A is identical to the characteristic function of a crisp set [50, 93]. Formally, the crisp set has the following property:

$$\mu_A(\theta) = \begin{cases} 1, & \text{iff } \theta \in A, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mu_A(\theta)$ of the crisp set is viewed as a characteristic function. Another convenient formulation for Eq. C.1 was proposed by Zadeh [87]. If Θ is a finite set $\{\theta_1, \theta_2, \ldots, \theta_n\}$, a fuzzy set A on Θ is expressed as

$$A = \mu_A(\theta_1)/\theta_1 + \ldots + \mu_A(\theta_n)/\theta_n = \sum_{i=1}^{n} \mu_A(\theta_i)/\theta_i \, . \tag{C.2}$$

If Θ is not finite, we write

$$A = \int_{\Theta} \mu_A(\theta)/\theta \, . \tag{C.3}$$

Equation C.3 is not in the conventional notation of integration. It is merely a continuous counterpart of the discrete case. In this sense, the integration representation remains valid when Θ is countable rather than a continuum [87].

Definition: Two fuzzy sets A and B are said to be equal (denotes $A = B$) iff

$$\forall \theta \in \Theta, \ \mu_A(\theta) = \mu_B(\theta) \, .$$

In crisp mathematics, there are only two ways to combine sets systematically, *i.e.*,

$$\theta \in A \ , \ \theta \in B \Rightarrow \theta \in A * B$$

$$\theta \in \overline{A} \ , \ \theta \in \overline{B} \Rightarrow \theta \notin A * B,$$

where $*$ must be either union ($\cup$) or intersection ($\cap$). This is no longer applicable when dealing with the fuzzy sets. Instead, other operators are used such as "min" and "max." These two operators are not the only ones to combine fuzzy sets. There are two other operators that do not have classical counterparts in traditional mathematics, such as the median operator, which lies between the "max" and "min" operators, and the symmetric sums, which are self–dual in the DeMorgan sense [73]. The behavior of these two operators is quite different from what we are familiar with. Note that elements with null membership can be omitted in Eq. C.2, and only elements with non–zero membership are considered in the fuzzy representation in Eq. C.1. Some numerical examples related to what has been discussed so far are presented below.

Example 2.3.1 [50]: If $A = \{\theta | \Theta$ is a real number and $\theta >> 1\}$, this is considered an ill–defined set. Subjectively, this set can be defined through a membership function, such as

$$\mu_A(\theta) = \begin{cases} 0, & \text{for } \theta \leq 1 \\ (\theta - 1)/\theta, & \text{for } \theta > 1 \end{cases}.$$

Despite the subjective definition of fuzzy sets, due to the nature of the problem that reflects its meaning and the way it is viewed, the membership function cannot be arbitrarily assigned [50]. For instance, we assign the membership function in Example 2.3.1 as

$$\mu_A(\theta) = \begin{cases} (\theta - 1)/\theta, & \text{for} \quad \theta \leq 1 \\ 0, & \text{for} \quad \theta > 1. \end{cases}$$

Another mistaken assignment would be if the membership function decreases monotonically as θ increases, for $\theta > 1$ such as

$$\mu_A(\theta) = \begin{cases} 0, & \text{for} \quad \theta \leq 1 \\ \exp -(\theta - 1), & \text{for} \quad \theta > 1. \end{cases}$$

Functions such as these are called nonadmissible functions of the fuzzy set A [50, 93]. In contrast, the following representation for Example 2.3.1 is called an admissible function of A if

$$\mu_A(\theta) = \begin{cases} 0, & \text{for } \theta \leq 1 \\ 1 - [\cosh(\theta - 1)]^{-1}, & \text{for } \theta > 1, \end{cases}$$

which satisfies the condition $0 \leq \mu_A(\theta) \leq 1$ for all $\theta \in \Theta$ and is consistent with the given example. Other important properties are the support, the height, and the crossover point of a fuzzy set. Such properties are best presented through examples.

Example 2.3.2 [50]: Let the universe be the interval, $\Theta = [0, 120]$, with θ interpreted as age. A fuzzy subset A of Θ labelled "old" may be defined as follows:

$$\mu_A(\theta) = \begin{cases} 0 & \text{for } 0 \leq \theta \leq 40 \\ [1 + [(\theta - 40)/5]^{-2}]^{-1} & \text{for } 40 \leq \theta \leq 120. \end{cases}$$

From the membership function assignment, the label, "old," is a sample within the interval $[40, 120]$; the height of "old" is the *supremum* of $\mu_A(\theta)$ over A, which is 1; and the *crossover point* of "old" is when the grade of membership function in A is 0.5, is 45.

Example 2.3.3 [50]: If the universe comprising Θ = {positive integers} and the fuzzy set, A, labeled "integers approximately equal to 5," A may be defined according to Eq. C.1 as

$$A = (2, .1), (3, .4), (4, .9), (5, 1), (6, .9), (7, .4), (8, .1).$$

Note that Θ in this example is a finite set.

Definition: A fuzzy set is convex iff

$$\mu_A(\lambda\theta_1 + (1-\lambda)\theta_2) \geq \min(\mu_A(\theta_1), \mu_A(\theta_2))$$

$$\theta_1, \theta_2 \in \Theta, \lambda \in [0,1]\,.$$

Similarly, a fuzzy set is concave iff

$$\mu_A(\lambda\theta_1 + (1-\lambda)\theta_2) \leq \max(\mu_A(\theta_1), \mu_A(\theta_2))\,.$$

C.2 Fuzzy Sets Operations

Definition [21]: The membership function, μ_C, of the intersection $C = A \cap B$ is point–wise defined for all $\theta \in \Theta$ by

$$\mu_C(\theta) \;=\; \min\{\mu_A(\theta), \mu_B(\theta)\}\,. \tag{C.4}$$

Definition [21]: The membership function, μ_D, of the union $D = A \cup B$ is point–wise defined for all $\theta \in \Theta$ by

$$\mu_D(\theta) \;=\; \max\{\mu_A(\theta), \mu_B(\theta)\}\,. \tag{C.5}$$

Definition [21]: The membership function of the complement of a fuzzy set A, $\mu_{\overline{A}}$, is defined by

$$\mu_{\overline{A}}(\theta) \;=\; 1 - \mu_A(\theta),\;\; \theta \in \Theta\,. \tag{C.6}$$

Example 2.3.5 [93]: Let the fuzzy sets A and B be as follows:

$$A \;=\; \{(1,.2),(2,.5),(3,.8),(4,1),(5,.7),(6,0.3)\} \text{ and}$$

$$B \;=\; \{(3,.2),(4,.4),(5,.6),(6,.8),(7,1),(8,1)\},$$

where the set of the available objects is $\Theta = \{1, 2, \cdots, 8\}$. Then the intersection $C = A \cap B$ according to Eq. C.4 is

$$C = \{(3, .2), (4, .4), (5, .6), (6, .3)\}.$$

The union $D = A \cup B$ according to Eq. C.5, is

$$D = \{(1, .2), (2, .5), (3, .8), (4, 1), (5, .7), (6, .8), (7, 1), (8, 1)\}.$$

The complement, $\overline{A}$, according to Eq. C.6, is

$$\overline{A} = \{(1, .8), (2, .5), (3, .2), (5, .3), (6, .7), (7, 1), (8, 1)\}.$$

Note that members with zero grade are not present in the fuzzy sets, but they must be considered in the operations as was shown in this example. Other useful operators in fuzzy sets, such as inclusion and the cardinality measures, are of great importance.

Definition [64]: If Θ is a finite set, the measure, $|A| = \sum_{\theta \in \Theta} \mu_A(\theta)$, is defined as the cardinality of a fuzzy set A. $||A|| = |A|/|\Theta|$ is called the relative cardinality.

Definition: A is said to be included in $B (A \subseteq B)$ iff

$$\forall \theta \in \Theta \,,\, \mu_A(\theta) \leq \mu_B(\theta). \tag{C.7}$$

Obviously, if $A = B$, then $A \subseteq B$ and $B \subseteq A$. The following example is the convex combination and the inclusion in the sense of Zadeh [84].

Example 2.3.6: Let A, B, and Λ be arbitrary fuzzy sets on Θ. The convex combination of A, B, and Λ is denoted by (A, B, Λ). Then using the definition of convexity, we can write

$$\forall \theta \in \Theta \,, \mu_{(A,B,\Lambda)}(\theta) \;\; = \;\; \mu_\Lambda(\theta)\mu_A(\theta) + (1 - \mu_\Lambda(\theta))\mu_B(\theta). \tag{C.8}$$

As a result of the convex combination of A, B, and Λ and the property of inclusion, we deduce

$$\forall \Lambda,\ A \cap B \subseteq (A, B, \Lambda) \subseteq A \cup B\,. \tag{C.9}$$

Useful references on this subject include [21, 22, 50, 64, 84, 85, 93].

C.3 A Possibility Distribution Concept

A possibility distribution can be described by restricted fuzzy subsets. Formally, let Θ be the universe of discourse, with the generic element of Θ denoted by θ. Furthermore, suppose a variable X can take values in the universe of discourse; *i.e.*,

$$X \;=\; \theta\,, \tag{C.10}$$

where $\theta \in \Theta$. Let A be a fuzzy subset of Θ that is characterized by the membership function μ_A. If A is a fuzzy restriction on X, then A acts like an elastic constraint on the values $\theta \in \Theta$ that may be assigned to X; *i.e.*,

$$X \;=\; \theta : \mu_A(\theta)\,, \tag{C.11}$$

where $\mu_A(\theta)$ is the degree to which the constraint represented by A is satisfied when $\theta \in \Theta$ is assigned to X [86]. Because we define A as a constraint that must be satisfied when the assignment takes place, then if we let $R(x)$ be the restricted constraint associated with X, we can write

$$R(x) \;=\; A\,. \tag{C.12}$$

This equation is called the relational assignment equation. As a result, Zadeh in 1977 stated the following definition:

Definition: Let A be a fuzzy subset of a universe of discourse Θ, which is characterized by its membership function, μ_A, with the degree of membership, $\mu_A(\theta)$, interpreted as the compatibility of θ with the concept of A. Let X be a variable taking values in Θ, and let A act as a fuzzy restriction, $R(x)$ associated with X. The proposition, $R(x) = A$, associates a possibility distribution Π_x, *i.e.*,

$$\Pi_x = R(X). \tag{C.13}$$

As a result, the possibility distribution function associated with X can be denoted by π_x and is defined to be numerically equal to the membership function A, *i.e.*,

$$\pi_x \stackrel{\text{def}}{=} \mu_A. \tag{C.14}$$

Thus, $\pi_x(\theta)$, the possibility that $X = \theta$, is postulated to be equal to $\mu_A(\theta)$. Note that from Eqs. C.13 and C.12, we deduce

$$\Pi_x = A. \tag{C.15}$$

Possibility and Fuzzy Sets Measures

Confidence Measures: Let θ be the generic elements of the universe of discourse Θ. If θ is an impossible event, then the confidence of an impossible event is $g(\theta) = 0$. In contrast, if θ is a sure event, then the confidence is $g(\theta) = 1$. Formally,

$$g(\emptyset) = 0 \text{ and } g(\Theta) = 1. \tag{C.16}$$

Sugeno [76] proposed an axiom that reflects the relation between the confidence of events and the inclusion property if it exists. Formally, let θ_1 and $\theta_2 \in \Theta$.

$$\text{if } \theta_1 \subseteq \theta_2 \Rightarrow g(\theta_1) \leq g(\theta_2). \tag{C.17}$$

This axiom says that if event θ_1 implies another event θ_2, then the confidence of θ_2 must be at least as much as the confidence in θ_1. This property is also known as a "fuzzy measure" by Sugeno [76].

Sugeno's J_λ Fuzzy Measures: Sugeno [76] introduced the so–called J_λ fuzzy measures, which are a family of parameterized functions from Θ to $[0,1]$ satisfying Sugeno's axiom given in Eq. C.17 such that $\forall\theta_1,\theta_2 \in \Theta$; if $\theta_1 \cap \theta_2 = \emptyset$, then

$$J_\lambda(\theta_1 \cup \theta_2) \;=\; J_\lambda(\theta_1) + J_\lambda(\theta_2) + \lambda J_\lambda(\theta_1)\cdot J_\lambda(\theta_2)\,, \tag{C.18}$$

where $-1 \leq \lambda \leq \infty$. Sugeno's J_λ fuzzy measures reduce to credibility and plausibility, but not necessity and possibility if λ is assigned as $\lambda \geq 0$ and $-1 < \lambda \leq 0$, respectively. Dubois and Prade [22] pointed out the isomorphism relationship between probability and J_λ fuzzy measures, *i.e.*, the relation between $J_\lambda(\theta)$ and $J_\lambda(\overline{\theta})$:

$$J_\lambda(\overline{\theta}) \;=\; \frac{1 - J_\lambda(\theta)}{1 + \lambda J_\lambda(\theta)}\,. \tag{C.19}$$

Possibility and Necessity Measures: From Sugeno's axiom in Eq. C.17, it follows that

$$\forall\theta_1,\theta_2 \subseteq \Theta\,,\; g(\theta_1 \cup \theta_2) \geq \max(g(\theta_1), g(\theta_2))\,, \tag{C.20}$$

$$g(\theta_1 \cap \theta_2) \leq \min(g(\theta_1), g(\theta_2))\,. \tag{C.21}$$

As a result, the limiting case of the confidence measure is the possibility measure [86], which can be readily expressed as follows:

$$\forall\theta_1,\theta_2 \subseteq \Theta\,,\; \Pi(\theta_1 \cup \theta_2) = \max(\Pi(\theta_1), \Pi(\theta_2))\,. \tag{C.22}$$

The term of possibility can be justified by the following argument.

Example 2.3.7: Suppose $\theta_1 \subseteq \Theta$ is an event considered certain. A function, Π, taking values in the interval $[0,1]$ and satisfying Sugeno's axiom, is easily defined by

$$\Pi(\theta_2) = \begin{cases} 1, & \text{if } \theta_1 \cap \theta_2 \neq \emptyset \\ 0, & \text{otherwise .} \end{cases} \tag{C.23}$$

This means θ_2 is possible based on the assumption of θ_1. A similar example would be if we let $\theta \subseteq \Theta$, and $\overline{\theta}$ is the complement of θ, then Eq. C.22 implies

$$\max(\Pi(\theta), \Pi(\overline{\theta})) = 1 , \tag{C.24}$$

which is always true that either θ or its complement is completely possible.

If Θ is finite, then it is natural that every singleton in Θ is assigned a possibility measure; *i.e.*,

$$\forall\theta,\ \pi(\theta) = \sup\{\pi(\omega)|\omega \in \theta\} , \tag{C.25}$$

where $\pi(\omega)$ is the possibility distribution associated with θ, and *sup* of $\pi(\omega)$ denotes the *supremum* or the *least upper bound* of $\pi(\omega)$ [86].

The other limiting case is to impose equality in Eq. C.21, which yields the necessity measures N. Considering the satisfaction of Sugeno's axiom, one can write

$$\forall\theta_1, \theta_2 \subseteq \Theta,\ N(\theta_1 \cap \theta_2) = \min(N(\theta_1), N(\theta_2)) . \tag{C.26}$$

The term of necessity can be justified by the following argument.

Example 2.3.8: Let N be a function that takes values in the interval $[0,1]$; then N

can be easily constructed if we have a sure event $\theta \subseteq \Theta$. Therefore,

$$N(\theta) = \begin{cases} 1, & \text{if } \theta \subseteq \Theta \\ 0, & \text{otherwise .} \end{cases} \tag{C.27}$$

When $N(\theta)$ is equal to 1, it means that θ is sure (necessarily true). From Eqs. C.26 and C.22, we deduce that a function N satisfies Eq. C.26 iff the function Π defined by

$$\forall\theta\,,\ \Pi(\theta) = 1 - N(\bar{\theta}), \tag{C.28}$$

is a possibility measure. The relation in Eq. C.28 is called the duality equation of the two measures. Furthermore, Eq. C.28 means that one can always construct the necessity measure by knowing the possibility measure; *i.e.*,

$$N(\theta) = \inf\{1 - \pi(\omega) | \omega \notin \theta\}\,, \tag{C.29}$$

where *inf* of a set, S, is the *infinum* or the *greatest lower bound* of S. Necessity measures satisfy the relation

$$\min(N(\theta), N(\bar{\theta})) = 0, \tag{C.30}$$

which prohibits the two contrary events to be necessary at the same time. From Eqs. C.28 and C.30, it is clear that

$$\forall\theta \subseteq \Theta,\ \Pi(\theta) \geq N(\theta), \tag{C.31}$$

which makes sense since an event becomes possible before becoming necessary. In addition, the following equations establish a better relation between possibility and necessity than the relation in Eq. C.31; *i.e.*,

$$N(\theta) > 0 \ \Rightarrow\ \Pi(\theta) = 1 \tag{C.32}$$

$$\Pi(\theta) < 1 \ \Rightarrow\ N(\theta) = 0. \tag{C.33}$$

Differences of Possibility and Probability Measures: From probability theory, the probability of a union between two events can be written in terms of their separate probabilities; *i.e.*,

$$\forall\theta_1, \forall\theta_2 \text{ with } \theta_1 \cap \theta_2 = \emptyset, p(\theta_1 \cup \theta_2) = p(\theta_1) + p(\theta_2)\,. \tag{C.34}$$

Also, the probability committed to an event and to its negation must add up to one; *i.e.*,

$$p(\theta) + p(\overline{\theta}) = 1\,. \tag{C.35}$$

In contrast, from Eqs. C.24 and C.30 we have

$$\max(\Pi(\theta), \Pi(\overline{\theta})) = 1$$

$$\min(N(\theta), N(\overline{\theta})) = 0\,.$$

Therefore,

$$N(\theta) + N(\overline{\theta}) \leq 1\,. \tag{C.36}$$

$$\Pi(\theta) + \Pi(\overline{\theta}) \geq 1\,. \tag{C.37}$$

The differences between probability and possibility are quite obvious from Eqs. C.35 through C.37.

APPENDIX D

Multi–sensory Fusion Using Uncertain Models

Statistics and decision theory play a major role in process control by providing decision–making tools. It seems very logical (*i.e.*, common sense) that one must think systematically to find the best decision which closely approximates or estimates the outcome of an experiment. Most statisticians, however, agree that similar events most likely occur due to similar circumstances. This belief encourages researchers and decision makers to compare past similar circumstances to allow them to obtain the most likely estimate, and then make the best decision in the present. It also makes sense if a group of information sources is combined to determine the decision criteria by which the decision utility is enhanced. Statistical assessments, however, may be different from one information source to another. The strength of a statistical assessment depends on the credibility of the information source. Statisticians readily feel comfortable changing their assessments in the light of someone else's opinion [29]. We agree that reconciliation of assessments is preferrable by most statisticians, but the problem of combining evidence fairly is not a straightforward one. To illustrate this difficulty, let us assume that you are the decision maker in a company. Suppose that you are unfamiliar with a certain type of problem, hence you would like to consult a panel of experts for advice. As a decision maker, it is legitimate to evaluate each expert's opinion based upon his/her expertise in a given problem. This crucial task arises due to the fact that experts are not alike and each one of them has his own experience, hence, his own judgment. By doing

this, you evaluate the suggested assessments, and you ought to come up with your own decision. If they all agree, you may adopt their consensus evaluation. Otherwise, you need to somehow process all assessments of your experts and come up with a substitutive judgment. From a mathematical point of view, it is extremely important to spell out the nature of the problem regarding the dependency on all pieces of information on one hand and the inter–dependency among all information sources on the other hand. The task is even more complex if the information sources are dependent. Combining evidence in a group decision framework can be done through probabilistic methods such as Bayes' theorem or through non–probabilistic methods even when the quantities in question are basic probability numbers. The latter are more general methods based on the recent development of the mathematical theory of evidence and the theory of fuzzy sets. In the subsequent sections, we will illustrate in more detail several types of problems related to combining evidence and the performance of each technique in both probabilistic and non–probabilistic methods. Before we explore combining evidence using different methods, we will first present an example that illustrates how we obtain a judgmental model for an unknown population.

D.1 Data Fusion Example

Consider the following dialogue that takes place between an analyst and his client regarding the uncertain statement, "young man." The analyst's goal is to construct a model that reflects a judgmental opinion of the client. Let us denote the uncertain statement, "young man," by a variable, θ, which can take any value from 0 to∞.

Analyst: I would like to show you that you can obtain a model that describes your judgmental opinion. I will also choose a context that is meaningful to you.

Client: O.K.

Analyst: In your opinion, how old is θ ?

Client: It is more likely that θ is 25 years old.

Analyst: Will you consider that θ is 25 years old to be the highest support ?

Client: Yes, I will support this assessment 100%.

Analyst: I understand that your estimate yields a separation between two regions for θ.

Client: Yes.

Analyst: Do you not think that the uncertain statement θ could mean a person whose age is between 25 and 30 years ?

Client: Yes, this is a possible assessment, but I will give it less support; say, between 90% and 100%.

Analyst: Will you give the same support for θ if θ is ranging from 20 to 25 years ?

Client: Yes.

Analyst: You do not have to be afraid of being inconsistent in your judgmental opinion. You should feel free to express your support without even thinking about past statements.

Client: Good. I feel more comfortable.

Analyst: What is the support that you will give if θ is ranging from 30 to 35 years old ?

Client: I will give less support than the previous assessments, that is, between 75% and 85%.

Analyst: Will you give the same support for θ ranging from 15 to 20 years old ?

Client: Yes.

Analyst: As I increase or decrease θ, your judgmental support decreases. Am I right ?

Client: Yes, but I may not decrease the support of θ in a symmetrical fashion.

Analyst: Good. What support will you assign if θ is ranging from 40 to 50 years old ?

Client: Very little, say between 10% and 30%.

Analyst: What support will you assign if θ is ranging from 5 to 10 years old ?

Client: I will say between 0% and 10%.

Analyst: What support will you assign for either $\theta > 50$ years old or $\theta < 5$ years old ?

Client: I will assign no support, that is, 0% for both of them.

Hence, one can easily construct a model that characterizes a judgmental opinion regarding the uncertain statement $\theta =$"young man." The constructed model is depicted in Fig. D.1. The example shows how a judgmental model is built by determining the opinion for each individual event. By fitting a smooth curve through these points and scaling, a continuous function of the judgmental opinion is obtained. This representation is assumed for clarity of the results. Consensus opinions for each fusion method given below are likewise shown as continuous. In all cases, however, data are combined using discrete values with the result plotted as a continuous function.

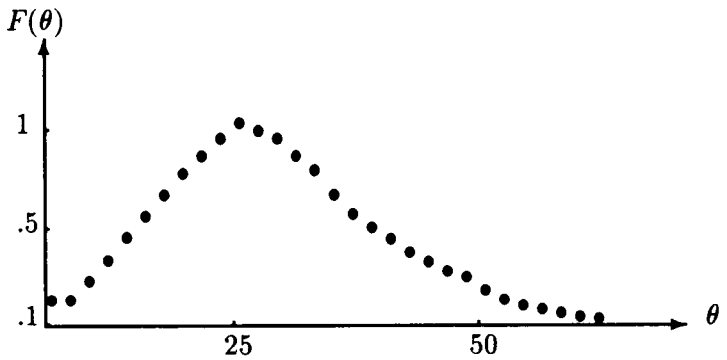


Figure D.1: *An uncertain model for the unknown θ.*

Let us assume that the analyst has conducted two separate experiments, $\mathcal{A}$ and $\mathcal{B}$. In each experiment, the analyst has consulted two different people by asking the same questions regarding the uncertain statement θ ="young man." In experiment $\mathcal{A}$, the analyst has found that his clients relatively agree with one another. In experiment $\mathcal{B}$, the clients disagree with one another. We may normalize the resulting models if necessary. Hence, each judgmental model reflects a personal opinion whose truth or falsity has yet to be determined. The results of both experiments $\mathcal{A}$ and $\mathcal{B}$ are illustrated in Fig. D.2. These two experiments will be applied to the probability, the evidence, and the possibility theory. Data fusion from each method will be compared with the others.

D.2 Combining Evidence Using the Probability Theory

In the literature, there are several techniques by which one can combine probabilistic evidence. Bayesians argue that many of the proposed methods are not always viable (*ad hoc* methods). However, they believe that the best approach to combining evidence is by conducting honest probabilistic modeling of the problem and by processing probabilistic

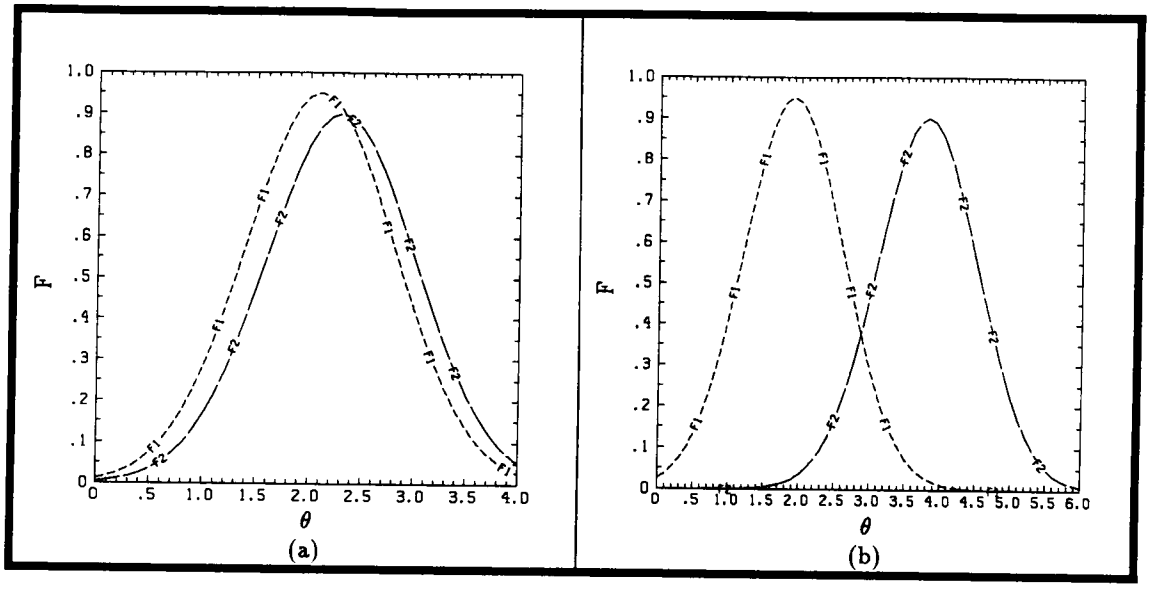


Figure D.2: *(a) judgmental models in relative agreement; (b) judgmental models in disagreement.*

inputs using the machinery of probability (of which Bayes' theorem is one aspect) [7]. From the data fusion example, the judgmental models in both experiments are considered as actual probability distributions; they may be normalized if necessary. In this section, we will apply these distributions to several techniques of combining evidence in the framework of probability theory. At the end of this section, a brief comparison among different aspects for combining probabilistic evidence is made.

Linear Opinion Pool

The linear opinion pool (L.O.P.) is one of the most widely known methods for combining evidence that appears in the literature [7, 62]. Users of this method believe that consensus distribution can be simply reached by taking the arithmetic mean of all pieces of information with the exception that the decision maker assigns different weights α_1, $\alpha_2, \ldots, \alpha_n$ to each information source. Using this concept, the consensus opinion, p_c,

over the same sample space, Θ, for any event θ in Θ is given by

$$p_c(\theta) = \sum_{i=1}^{n} \alpha_i p_i(\theta), \tag{D.1}$$

where

$$\sum_{i=1}^{n} \alpha_i = 1.$$

Therefore, one must determine the weight parameter from the strength of different opinion distributions. To do so, one may suggest using the following method. To make the discussion simpler, assume that $n = 2$; then Eq. D.1 becomes

$$p_c(\theta) = \alpha_1 p_1(\theta) + \alpha_2 p_2(\theta), \tag{D.2}$$

where

$$\alpha_1 + \alpha_2 = 1.$$

Let $\alpha_1 = \alpha$, then $\alpha_2 = 1 - \alpha$. Substituting in Eq. D.2, we get

$$p_c(\theta) = \alpha p_1(\theta) + (1 - \alpha) p_2(\theta). \tag{D.3}$$

Now, our goal is to determine α in order to assign the right weight for each given assessment. One way to do this is to seek α that minimizes

$$\int_{\Theta} [\alpha p_1(\theta) + (1 - \alpha) p_2(\theta) - 1/2]^2 \, d\theta.$$

Note that 1/2 represents the ignorance as it was defined by the Bayes' rule. The last equation can be written in a simpler form; *i.e.*,

$$\int_{\Theta} (A\alpha + B)^2 d\theta = 0, \tag{D.4}$$

where $A = p_1(\theta) - p_2(\theta)$ and $B = p_2(\theta) - 1/2$. Further simplifications yield

$$\alpha^2 \int_\Theta A^2 d\theta + 2\alpha \int_\Theta AB d\theta + \int_\Theta B d\theta = 0.$$

Or

$$K_1\alpha^2 + 2K_2\alpha + K_3 = 0,$$

where

$$\begin{aligned} K_1 &= \int_\Theta A^2 d\theta, \\ K_2 &= \int_\Theta AB d\theta, \\ K_3 &= \int_\Theta B^2 d\theta. \end{aligned} \tag{D.5}$$

Solving for α, we get

$$\alpha = \frac{-K_2 \pm \sqrt{K_2^2 - K_1 K_3}}{K_1}. \tag{D.6}$$

Then, a consensus distribution, p_c, for only two different assessments can be found by using Eqs. D.3 through D.6.

The linear opinion pool seems quite promising if the decision maker can precisely assign the right weight to each information source. Unfortunately, this is not always possible. Practically, the real test of a method's performance is in its effectiveness in dealing with contradicting bodies of evidence. To show the effect of both supporting and conflicting bodies of evidence, the example developed in Section 3.1 was applied to the L.O.P., as shown in Fig. D.3. Two different cases are applied. First, the information sources are in relative agreement. Second, the information sources are in disagreement. For the

test cases in relative agreement, Fig. D.3(a) curve P3 shows that the L.O.P. consensus opinion has a value which is close to and between the two sources shown as curves P1 and P2. For the relative disagreement case, Fig. D.3(b) curve P3 indicates that the consensus will be closest to the source that is significantly larger than the other. When the two sources have equal probability, the result is that probability value.

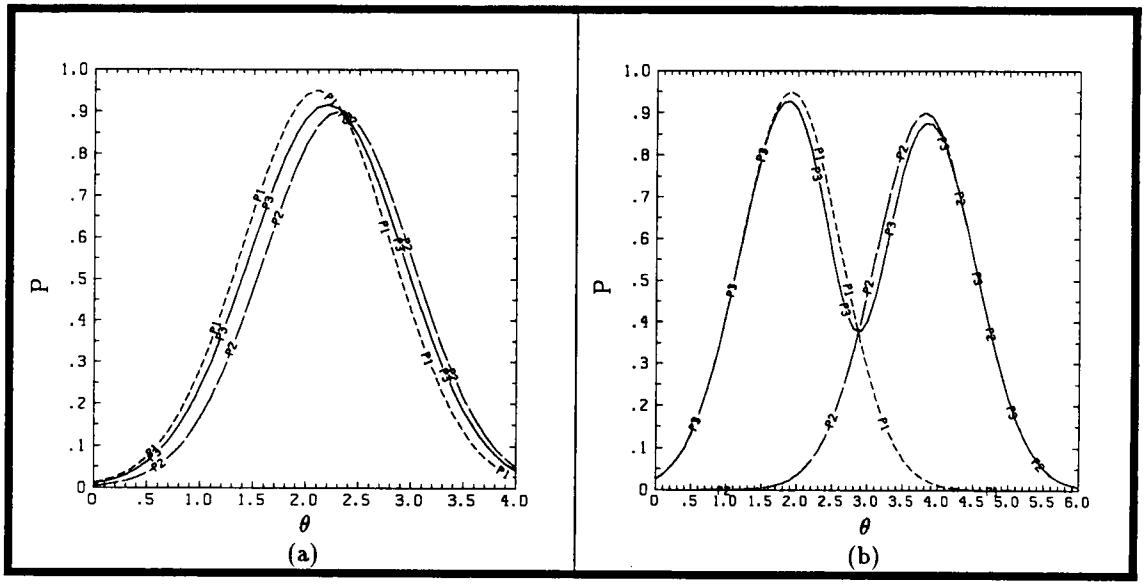


Figure D.3: *(a) L.O.P. method for experiment $\mathcal{A}$; (b) L.O.P. method for experiment $\mathcal{B}$.*

Independent Opinion Pool

The independent opinion pool (I.O.P.) is a well–known method for averaging statistical quantities. It simply computes the geometric mean, which is considered as the consensus opinion, of given statistical opinions. However, the use of this method must be handled with great care. Generally, the consensus $p_c(\theta)$ over Θ for an event θ using this method is given by

$$p_c(\theta) = K\left[\prod_{i=1}^{n} p_i^{\alpha_i}(\theta)\right], \tag{D.7}$$

where K is an appropriate normalizing constant, n is the number of independent opinions, and the α_i are positive constants. The latter is used to satisfy the basic properties of the probability theory, and it is given by

$$\sum_{i=1}^{n} \alpha_i = 1. \tag{D.8}$$

The independent opinion pool is not always valid for combining statistical quantities because that method puts restrictions on the type of information sources that can be consulted; *i.e.*, the information sources must be independent in order to validate the use of them. At any rate, it is quite adequate to use the independent opinion pool in a case such as combining assessments reported by a panel of experts, where each has his own experience and subsequently his own judgment. Practically, this special case is hard to find when one consults two or more computer sets for advice. For instance, if one of the information sources has a basic probability number of zero, it allows one of the information sources to dictate the resulting consensus, which will also be zero. The other sources are reinforced to have no opinion at all. The two test cases from the example are applied to the I.O.P. models with the results shown in Fig. D.4. The I.O.P. method has a consensus, which is near to and between the two sources in relative agreement as shown in Fig. D.4(a) curve P3. This result is similar to the L.O.P. method. The consensus for the sources in relative disagreement as given in Fig. D.4(b), however, is closer to the smaller of the sources when there are significant differences. When the two sources have equal probability, the consensus is that same value just as in the L.O.P. case.

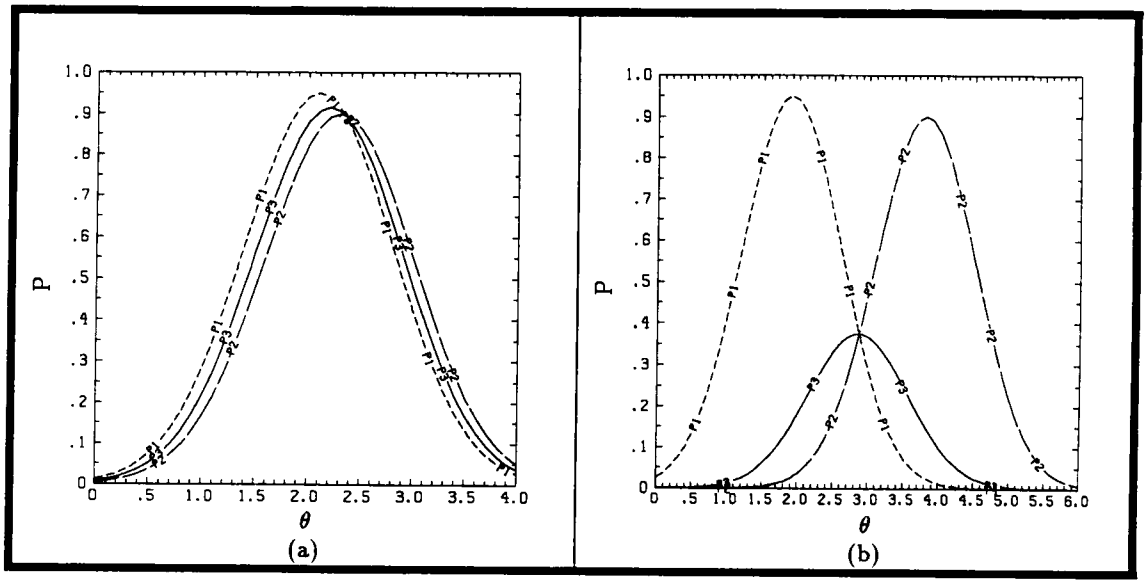


Figure D.4: *(a) I.O.P. method for experiment $\mathcal{A}$; (b) I.O.P. method for experiment $\mathcal{B}$.*

The Harmonic Opinion Pool

The harmonic opinion pool (H.O.P.) is another type of method for averaging several pieces of information. It computes the reciprocal of the mean for reciprocals:

$$p_c(\theta) = \frac{n}{\sum_{i=1}^{n} \frac{1}{p_i(\theta)}}, \tag{D.9}$$

where $p_c(\theta)$ is the consensus opinion over Θ for an event θ and n is the number of the suggested opinions. Clearly, this method shares the inadequancies of the I.O.P. One of these inadequancies is that it cannot handle the situation where at least one of the suggested basic probability numbers is zero. Figure D.5 gives the consensus for the H.O.P. method after applying the two example test cases. In Fig. D.5(a) curve P3 shows that the relative agreement consensus is between and near the two sources, very similar to the L.O.P. and I.O.P. methods. Consensus for the relative disagreement case in Fig. D.5(b) tends to follow the smaller of the two sources when significant differences

are present, while the value is the same when the sources are equal. This behavior is similar to the I.O.P. except that the consensus in the H.O.P. is even closer to the smaller of the sources for large differences.

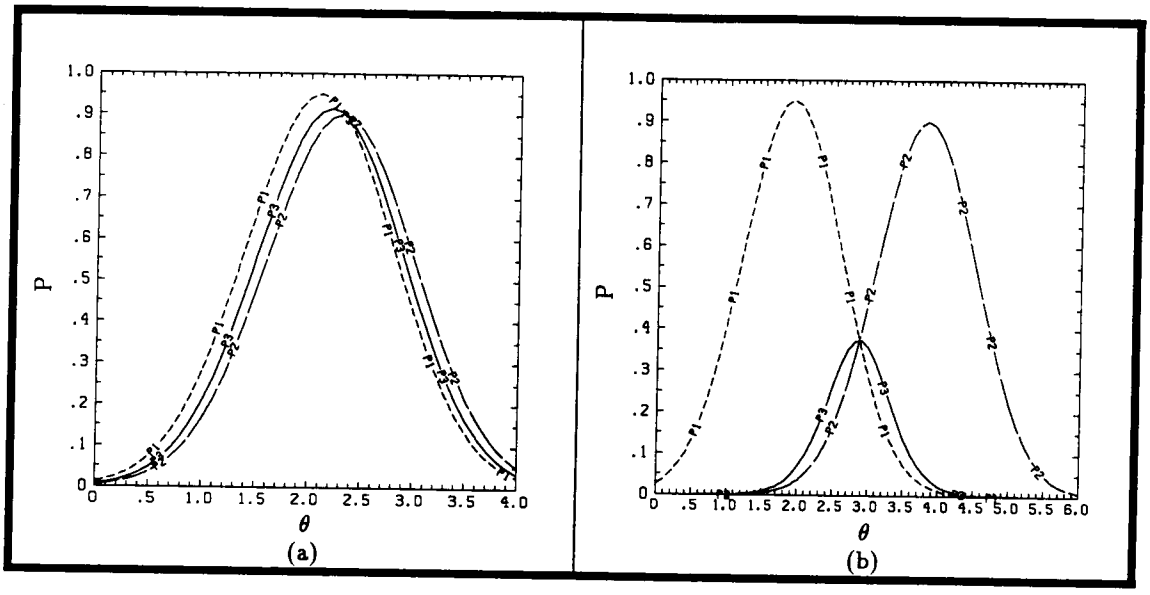


Figure D.5: *(a) H.O.P. method for experiment A; (b) H.O.P. method for experiment B.*

The Super Bayesian Approach

The super Bayesian approach is a direct application of Bayes' theorem [7]:

$$p(\theta_i|x) = \frac{p(x|\theta_i) \cdot p(\theta_i)}{\sum_{j=1}^{n} p(x|\theta_j) \cdot p(\theta_j)},$$

where $i = 1, 2, \ldots, n$, so that for every subscript of θ, we would have a related posterior probability associated with it. Thus, one can compute the revised assessment concerning the occurrence of x. To derive the Super Bayesian Approach, let us assume that two sources of information, π_1 and π_2, are consulted, and their opinions regarding the occurrence of the event, θ_1, are $\pi_1(\theta_1)$ and $\pi_2(\theta_1)$, respectively. Let us assume further

that the occurrence of θ_1 will be the complement of the occurrence of θ_2. Then, one can write the assessment of the first information source π_1 as follows:

$$\pi_1(\theta_1) = p, \tag{D.10}$$

$$\pi_1(\theta_2) = 1 - p. \tag{D.11}$$

Note that each source of information is assumed to be coherent. When using Bayes' theorem, one of the information sources can be specified as the decision maker, which is the *a priori* probability. The other information sources are considered likelihood functions or suggested assessments. In this case, assume that the decision maker has *a priori* probability $\pi_2(\theta_1)$. The problem is to process the assumed *a priori* probability and the suggested likelihood to find the final estimate of the occurrence of θ_1. To do so, we must find the densities $\mathcal{F}(p|\theta_1)$ and $\mathcal{F}(p|\theta_2)$ for the information sources' opinions in past similar situations. These may be found by running an experiment and recording the most "likely" values of occurrences. Once $\mathcal{F}(p|\theta_1)$ and $\mathcal{F}(p|\theta_2)$ have been specified, Bayes' theorem can be applied as follows

$$\pi(\theta_1|p) = \frac{\mathcal{F}(p|\theta_1) \cdot \pi_2(\theta_1)}{\mathcal{F}(p|\theta_1) \cdot \pi_2(\theta_1) + \mathcal{F}(p|\theta_2) \cdot \pi_2(\theta_2)},$$

or

$$\pi(\theta_1|p) = \left[1 + \frac{\mathcal{F}(p|\theta_2)}{\mathcal{F}(p|\theta_1)} \cdot \frac{\pi_2(\theta_2)}{\pi_2(\theta_1)}\right]^{-1}. \tag{D.12}$$

If the functions, $\mathcal{F}(p|\theta_1)$ and $\mathcal{F}(p|\theta_2)$, are not available, and we know that the occurrence of θ_1 will be the complement of the occurrence of θ_2, then reasonable distributions with roughly the implied means and variances are the $\mathcal{B}(2,1)$ and $\mathcal{B}(1,2)$ distributions, respectively, where $\mathcal{B}(x,y)$ is Beta. Therefore, it might be acceptable to specify a mean (m) and standard deviation (δ) for the source's opinion (p) when θ_1 is true and then

model this by choosing $\mathcal{F}(p|\theta_1)$ to be a Beta density. The Beta distribution function can be specified as follows:

$$\mathcal{B}(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \text{ for } \alpha > 0 \text{ and } \beta > 0. \tag{D.13}$$

It is related to the Gamma function through

$$\mathcal{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{D.14}$$

If the integral of Eq. D.13 is divided by the expression in the Eq. D.14, then we get

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx = 1\,,\ 0 < x < 1\,.$$

By definition, if

$$\int_0^1 \mathcal{F}(x)dx = 1,$$

then $\mathcal{F}(x)$ is the Beta density function. In our example, if we consider $\mathcal{F}(p|\theta_1)$ to be the Beta function $\mathcal{B}(\alpha,\beta)$, and $\mathcal{F}(p|\theta_2)$ to be the Beta function $\mathcal{B}(\beta,\alpha)$, then we can write

$$\mathcal{F}(p|\theta_1) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \tag{D.15}$$

$$\mathcal{F}(p|\theta_2) = \frac{\Gamma(\beta+\alpha)}{\Gamma(\beta)\Gamma(\alpha)} p^{\beta-1}(1-p)^{\alpha-1}. \tag{D.16}$$

Substituting Eqs. D.15 and D.16 into Eq. D.12, we get

$$\pi(\theta_1|p) = \left[1 + \left(\frac{1-p}{p}\right)^{\alpha-\beta} \cdot \frac{\pi_2(\theta_2)}{\pi_2(\theta_1)}\right]^{-1}. \tag{D.17}$$

Suppose the decision maker's opinion about the occurrence of θ_1 (*a priori* probability) is p_2, then the decision maker's opinion about the occurrence of θ_2 is $(1-p_2)$, which is *a priori* probability, too. To illustrate this, let p be equal to p_1, and assume further that $\alpha = 2$ and $\beta = 1$ [7]. Then, Eq. D.17 becomes

$$\pi(\theta_1|p) = \left[1 + \frac{(1-p_1)}{p_1} \cdot \frac{(1-p_2)}{p_2}\right]^{-1},$$

$$\pi(\theta_1|p) \;=\; \frac{p_1 p_2}{p_1 p_2 + (1-p_1)(1-p_2)}. \qquad \text{(D.18)}$$

This is known as the Super Bayesian Approach for combining evidence and group decisions [7, 27].

Having derived the Super Bayesian Approach of combining evidence and group decisions, it would be appropriate to discuss its practical validity. The following weaknesses are present: (1) Most statisticians agree that the Super Bayesian Approach does not offer fair treatment between the old evidence and the new evidence. It considers the new evidence a proposition by which the old evidence (*a priori* probability) is conditioned. In other words, the new evidence sometimes will be facing an unfair decision that is made by the *a priori* probability (the decision maker). There is no symmetry of evaluation between the evidence and group decisions. (2) Bayesians pretend to have detailed prior opinions with the proposition being a subset of a set over which priors are defined. (3) Bayesians are unable to represent ignorance. This last statement can be best illustrated through an example. Let us assume that the set Θ has only two events, θ_1 and θ_2, where θ_2 is the complement of θ_1: $\Theta = \{\theta_1, \theta_2\}$. According to Bayesians, $Bel(\theta_1) = 1/2$ and $Bel(\theta_2) = 1/2$ if there is no information available about θ_1 and θ_2, where $Bel(\theta_1)$ and $Bel(\theta_2)$ are the belief functions committed to θ_1 and θ_2 respectively; *i.e.*,

$$Bel(\theta_1) + Bel(\theta_2) \;=\; 1\,. \qquad \text{(D.19)}$$

Let us assume further that θ_1 contains two distinct pieces of information, say λ_1 and λ_2. Now, Bayesians must adjust their representation of ignorance. Therefore, it is necessary that a new set be introduced. Let us assume the new set is $\Omega = \{\lambda_1, \lambda_2, \lambda_3\}$, where λ_3 is equivalent to θ_2. Then, the representation of ignorance must be updated; *i.e.*,

$Bel(\lambda_1) + Bel(\lambda_2) + Bel(\lambda_3) = 1$, and $Bel(\lambda_1) = Bel(\lambda_2) = Bel(\lambda_3) = 1/3$. From the latter equation we see that $Bel(\lambda_1)+Bel(\lambda_2) = 2/3$ and $Bel(\lambda_3) = 1/3$, which contradict the representation of $Bel(\theta_1)$ and $Bel(\theta_2)$. Therefore, Bayes' theorem is inconsistent in representing ignorance. In other words, it cannot distinguish between lack of belief and disbelief [67]. (4) Bayesians assume that the assessment makers are coherent. This could be true if a person judges a particular event. It is, however, far from reality if a machine makes the judgment. Quite obviously the reason is that machines make their judgments by reasoning about circumstances and events, and they do not allow past assessments to influence their process in making new ones.

The performance of the super Bayesian approach for the example test cases is shown in Fig. D.6. In the relative agreement case, Fig. D.6(a), when the sources have a relatively high value, the consensus opinion is larger than either of the two sources and reaches a value near one over a range of the largest source values. However, the consensus is smaller than either of the sources for small source values and approaches zero quickly as the source values decrease.

For the sources in relative disagreement, the super Bayesian consensus is given in Fig. D.6(b). Note the asymmetry of the consensus curve P3, which has a higher peak value for source curve P1 than for P2, indicating the non-equal consideration of the old and new evidence. The consensus is always smaller than the larger source, and approaches zero when one of the sources approaches zero regardless of the value of the other source.

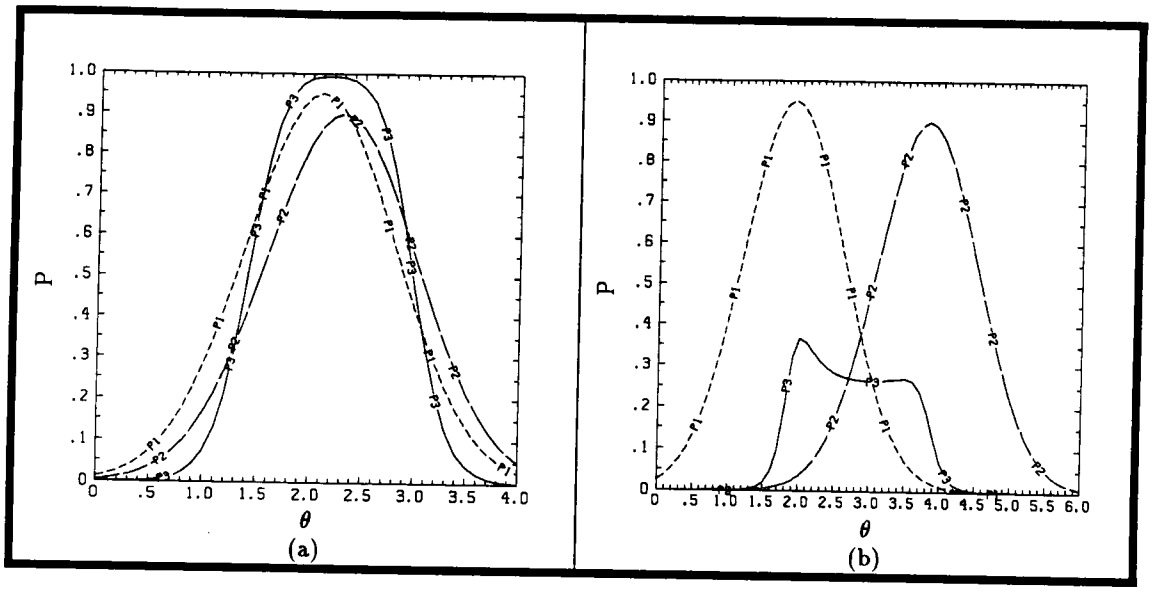


Figure D.6: *(a) Super Bayesian Approach for experiment A; (b) Super Bayesian Approach for experiment B.*

D.3 Combining Evidence Using the Evidence Theory

In this method, evidential reasoning is based upon the Dempster–Shafer theory of belief functions, which is considered more general than the techniques found in probability theory [67]. Dempster's rule of combination is the heart of the theory of evidence (belief). Using the data fusion example in Section 3.1, degrees of support are considered as basic probability numbers. In this framework, each source of information commits partial belief toward a proposition (*i.e.*, a focal element). Here, we represent the data fusion example as a collection of several propositions, each of which is supported by the two knowledge sources m_1 and m_2 to the degrees S_1 and S_2, respectively. The degrees of support, S_1 and S_2, are ordinary probability assignments, which can take any value between 0 and 1. To apply the theory of belief, we must first present Dempster's rule of combination, which enables us to compute the combined evidence.

Dempster's Rule of Combination

Two or more belief functions can be combined using Dempster's rule of combination. Specifically, Dempster's rule of combination finds a new belief function based on the given belief functions from different knowledge sources by computing the orthogonal sum of those belief functions. Usually, we express belief functions in terms of their basic probability numbers. Let Bel_1 and Bel_2 be the belief functions over a frame of discernment Θ, and denote Bel_1's and Bel_2's focal elements by $A_1, A_2, \cdots, A_k$, and $B_1, B_2, \ldots, B_\ell$, respectively. Then, the basic probability numbers that are suggested by Bel_1 and Bel_2 are $m_1(A_1), m_1(A_2), \ldots, m_1(A_k)$, and $m_2(B_1), m_2(B_2), \ldots, m_1(B_\ell)$, respectively. A geometric representation can be formed for both belief functions, as illustrated in Fig. D.7.

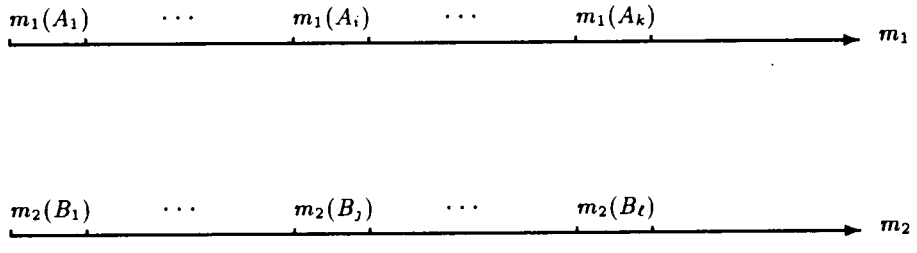


Figure D.7: *Probability assignment for both knowledge sources.*

Basic probability numbers have been assigned to a set of propositions, from which the evidential interval can be determined.

I. A Single Belief Function:

Part of the total belief that is committed to a proposition, "focal element," can be

stated as follows. Consider a frame of discernment Θ. A function m: $2^{\Theta} \rightarrow [0,1]$ is called a basic probability number whenever

1. $m(\emptyset) = 0$

2. $\sum_{A \subseteq \Theta} m(A) = 1.$

Using this definition and the geometric representation for the belief function shown in Fig. D.7, we see that one's belief can be partitioned among all subsets of the frame of discernment Θ such that $m(A)$ is the quantity of A's basic probability number. The lower probability, $Cr(A)$, "credibility," represents the minimum amount of belief that would be committed to a proposition A. The latter is often called the degree of support for a proposition A, which is given by

$$Cr(A) = S(A) = \sum_{B \subseteq A} m(B).$$

Applying this definition to the geometric representation that we have in Fig. D.7, we get

$$S_1(A) = \sum_{A_i \subseteq A} m_1(A_i), \tag{D.20}$$

and the upper probability "plausibility" committed to A is

$$P\ell_1(A) = 1 - S_1(\overline{A}). \tag{D.21}$$

II. A Simple Support Function:

Let Bel: $2^{\Theta} \rightarrow [0,1]$, a belief function, $Bel(B)$, is called a simple support function, which supports the belief function, $Bel(A)$, to the degree S if

$$Bel(B) = \begin{cases} 0, & \text{if } B \not\supset A \\ S, & \text{if } B \supset A,\ B \neq \Theta \\ 1, & \text{if } B = \Theta. \end{cases} \tag{D.22}$$

The simple support function reflects the strength of evidence; it can be used to assign weights for the belief functions [67].

III. Composition of Belief Functions:

The geometric representation for both m_1 and m_2 completely satisfies the definition of belief functions. Therefore, the orthogonal sum of the two belief functions committed to $A_i \cap B_j$ can be determined by computing the probability mass of measure $m_1(A_i)m_2(B_j)$ committed to $A_i \cap B_j$. The orthogonal sum can be geometrically computed with the aid of Fig. D.8.

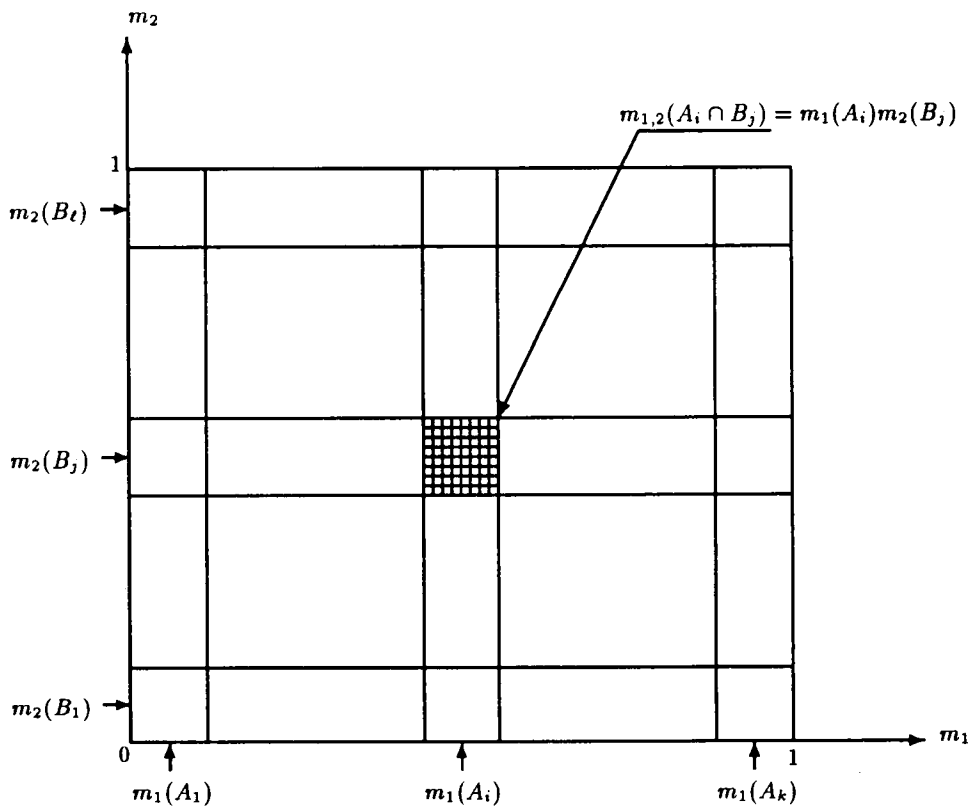


Figure D.8: *Composition of probability from both knowledge sources (Dempster's rule of combination).*

The dark rectangle represents the orthogonal sum of the two belief functions com-

mitted to $A_i \cap B_j$. Formally, a subset of Θ may have more than one rectangle exactly committed to it; the total probability mass committed to A, such as $A \subset \Theta$ and $A = A_i \cap B_j$, is

$$Bel(A) = \sum_{\substack{i,j \\ A_i \cap B_j = A}} m_1(A_i) m_2(B_j) . \tag{D.23}$$

This form represents the orthogonal sum of two belief functions, which is committed to $A = A_i \cap B_j$. However, we may face a problem of applying this formula if there is an empty set such as $A_i \cap B_j = \emptyset$. In this case,

$$\sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) > 0.$$

To overcome this problem, the orthogonal sum committed to an empty set such as $A_i \cap B_j = \emptyset$ must be subtracted from the total orthogonal sum committed to $A_i \cap B_j$. Further, adjustment to this formula satisfies the definition that the total belief committed to the space of discernment Θ is $\sum_{A \subseteq \Theta} m(A) = 1$. Therefore, the measures of all sub–orthogonal sums committed to $A_i \cap B_j = A$ must be multiplied by a normalization factor:

$$\left[1 - \sum_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) m_2(B_j) \right]^{-1} .$$

Hence, Dempster's rule of combination is

$$m(A) = \frac{\sum_{\substack{i,j \\ A_i \cap B_j = A}} m_1(A_i) m_2(B_j)}{1 - \sum_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) m_2(B_j)} . \tag{D.24}$$

For combining several belief functions, the process is repeated in a pairwise fashion. That is, the orthogonal sum is found for two belief functions producing a new belief function which is a candidate to be combined with a third. This process continues until all belief functions are considered.

Special Cases of Dempster's Rule of Combination

I. Bernoulli's Rule of Combination:

Suppose both information sources support $A \subset \Theta$ to a certain degree; *i.e.*, the first information source supports $A \subset \Theta$ to the degree S_1, and the second information source supports $A \subset \Theta$ to the degree of S_2. Bernoulli's rule of combination focuses on combining two support functions, S_1 and S_2, both of which support $A \subset \Theta$. Bernoulli argues that the total degree of support of A must be greater than any degree of support acting alone. Bernoulli proposed the following rule of combination:

$$S(A) = 1 - (1 - S_1)(1 - S_2), \tag{D.25}$$

where $S(A)$ is the total support committed for $A \subset \Theta$ by both sources of information. Bernoulli's rule of combination is a special case of Dempster's rule of combination. For this special case, the belief function, Bel_1, produces a basic probability number, $m_1(A)$, supporting A, and the belief function, Bel_2, produces a basic probability number, $m_2(A)$, supporting A; *i.e.*, $m_1(A) = S_1$ and $m_2(A) = S_2$. The geometric representation for both belief functions by which the orthogonal sum can be easily found is depicted in Fig. D.9.

From Fig. D.9, we have $m_1(\Theta) = (1 - S_1)$, and $m_2(\Theta) = (1 - S_2)$. Substituting in Eq. D.24, we get

$$m(A) = \frac{1 - m_1(\Theta) \cdot m_2(\Theta)}{1 - 0},$$

where $1 - m_1(\Theta) \cdot m_2(\Theta)$ is the belief committed to A. Note that there is no belief committed to the empty set. Thus,

$$S(A) = 1 - (1 - S_1)(1 - S_2), \tag{D.26}$$

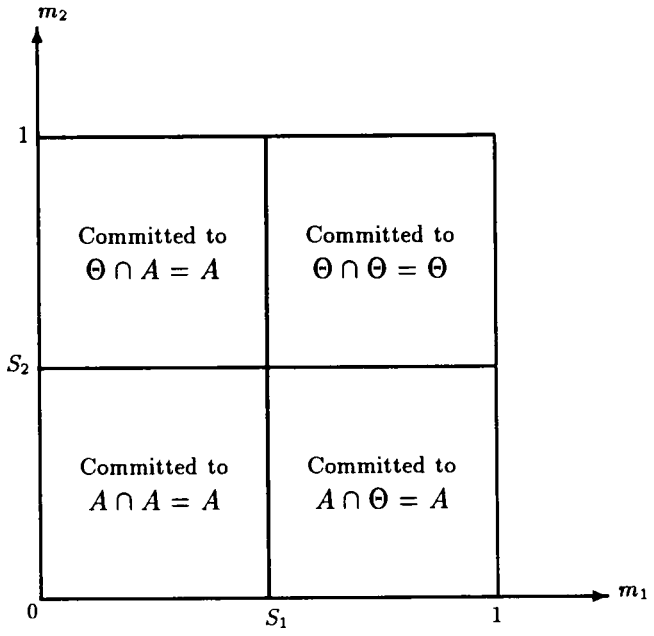


Figure D.9: *Composition of probability from both knowledge sources committed to the same event (Bernoulli's rule of combination).*

where $S(A) = m(A)$. Equation D.26 is the same as Eq. D.25. Hence, Bernoulli's rule of combination is a special case of Dempster's. One important fact we may notice is that the total area of the square is always one (see Fig. D.9).

II. Rules for Combining Heterogeneous Evidence with $A \cap B \neq \emptyset$:

This is another special case of Dempster's rule. In this case we have two belief functions, Bel_1 and Bel_2, committed to A and B, respectively. Formally, the basic probability numbers are

$$\begin{aligned} S_1(A) &= S_1 \\ S_2(B) &= S_2, \\ A \cap B &\neq \emptyset. \end{aligned}$$

The geometric representation of both belief functions is depicted in Fig. D.10.

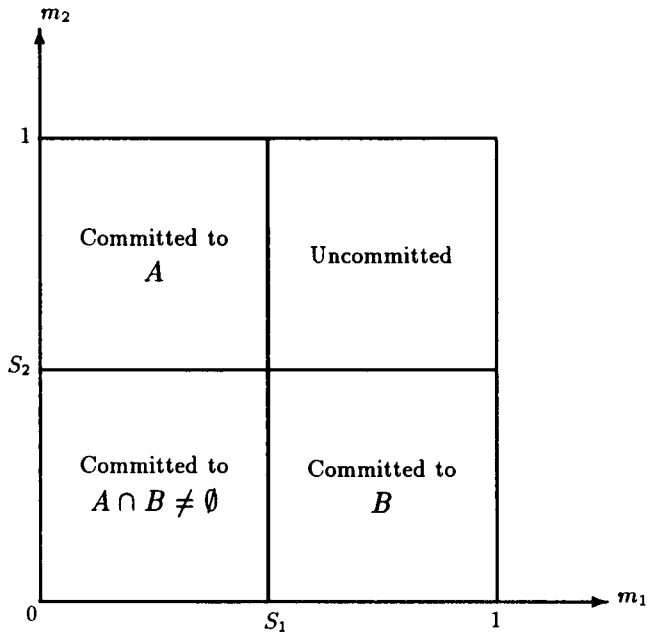


Figure D.10: *Composition probability of heterogeneous evidence for the case of* $A \cap B \neq \emptyset$.

From Fig. D.10, we compute the orthogonal sum for each square cell:

$$\begin{aligned} S(A \cap B) &= S_1 S_2 \\ S(A) &= S_1(1 - S_2) \\ S(B) &= S_2(1 - S_1) . \end{aligned} \tag{D.27}$$

Note that the support functions, S_1 and S_2, committed to A and B influence one another since $A \cap B \neq \emptyset$.

III. Rule for Combining Conflicting Evidence with $A \cap B = \emptyset$:

Let S_1 and S_2 be the support functions focused on A and B, respectively. Dempster's geometric representation can be illustrated as shown in Fig. D.11.

Because A does not intersect with B, substitution in Eq. D.24 yields

$$m(A) = \frac{S_1(1 - S_2)}{1 - S_1 S_2} , \tag{D.28}$$

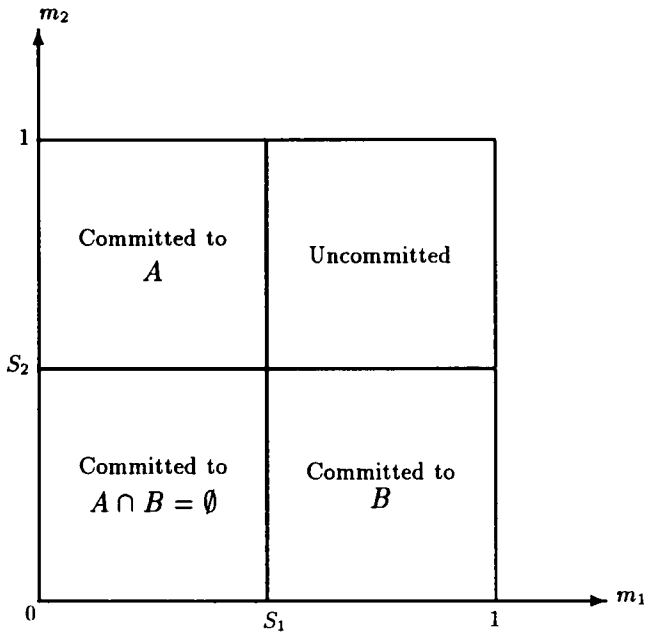


Figure D.11: *Composition probability for conflicting evidence ($A \cap B = \emptyset$).*

$$m(B) = \frac{S_2(1 - S_1)}{1 - S_1 S_2}, \tag{D.29}$$

$$m(\Theta) = \frac{(1 - S_1)(1 - S_2)}{1 - S_1 S_2}. \tag{D.30}$$

The Weight of Evidence

Let A be a subset in the space of discernment Θ. Assume that A has the degree of support, S, focused on A. Assume further that S has the weight of evidence ω. Two conditions must be considered:

1. $0 \leq \omega \leq \infty$

2. $0 \leq S \leq 1$.

It is sensible to consider that the support, S, is a function of the weight of evidence (*i.e.*, $S = g(\omega)$). Consider, for instance, Bernoulli's case where two support functions, S_1 and

S_2, are focused on A, with the weight of evidence ω_1 and ω_2, respectively. Hence, we require

$$g : [0, \infty] \rightarrow [0, 1]$$

$$g : S \rightarrow g(\omega) .$$

Since weights combine additively, the weight of the orthogonal sum, $S = S_1 \oplus S_2$, is $\omega_1 + \omega_2$. Hence,

$$g(\omega_1) = S_1 ,$$

$$g(\omega_2) = S_2 ,$$

and

$$g(\omega_1 + \omega_2) = 1 - (1 - S_1)(1 - S_2) . \tag{D.31}$$

If we set $F(\omega) = 1 - g(\omega)$, we can rewrite Eq. D.31 as follows:

$$F(\omega_1) = 1 - S_1 ,$$

$$F(\omega_2) = 1 - S_2 ,$$

and

$$g(\omega_1 + \omega_2) = 1 - F(\omega_1)F(\omega_2) . \tag{D.32}$$

From Eq. D.32 we can write

$$F(\omega_1 + \omega_2) = F(\omega_1)F(\omega_2) . \tag{D.33}$$

If $F(\omega)$ is continuous, either $F(\omega) \equiv 0$ or else

$$F(\omega) = e^{c\omega}$$

$$g(\omega) = 1 - e^{c\omega},$$

where c is an arbitrary constant. In order to map $[0, \infty]$ onto $[0, 1]$, the constant c must be negative. Setting $c = -1$, we get the relation between the weight of evidence ω and the support function S as follows:

$$\begin{aligned} S &= 1 - e^{-\omega} \\ \omega &= -\log(1 - S). \end{aligned} \tag{D.34}$$

In summary, if $E(\alpha)E(\beta) = E(\alpha + \beta)$, then $E(\alpha)$ and $E(\beta)$ are exponential.

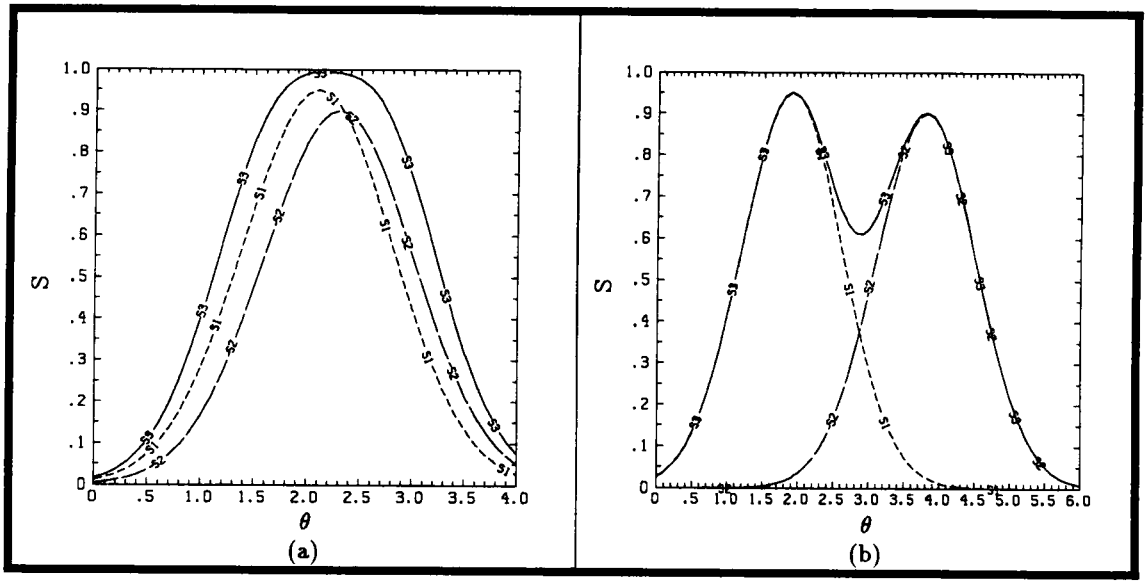


Figure D.12: *(a) Dempster's rule of combination for experiment $\mathcal{A}$; (b) Dempster's rule of combination for experiment $\mathcal{B}$.*

Using the two judgmental models taken from the example, aggregation based on Dempster's rule of combination and its special case (Bernoulli's rule of combination) is illustrated in Fig. D.12. The consensus opinion for Dempster's rule of combination is applied to the example shown in Fig. D.12. For the relative agreement case given in Fig. D.12(a), the consensus in curve S3 is larger than either of the two source cases

(curves S1, S2) and approaches one near the maximum values of the sources. The consensus follows but remains larger than the sources for the entire source range. This behavior is also seen in the relative disagreement case shown in Fig. D.12(b). When one source is near zero, the consensus follows the other source. However, the consensus becomes larger than either of the sources in the middle region where intermediate values are present.

D.4 Combining Evidence Using the Possibility Theory

We have shown how the possibility theory emerged from the theory of fuzzy sets, then stated differences and similarities between them. Nevertheless, the possibility theory obeys, of course, all axioms and definitions by which the theory of fuzzy sets was first established. Since Zadeh [84] stated the basic definitions of the theory of fuzzy sets, there have been conflicting opinions regarding the algebraic operators used in the combination processes of fuzzy sets. Such operators are union, intersection, and complementation. From a mathematical point of view, the dispute is perfectly valid because fuzzy sets are different from crisp sets, and in dealing with such new sets, it is more suitable to employ a different type of mathematics. This is known as fuzzy mathematics. In the context of fuzzy mathematics, operators have been found to be equivalent to intersection and union in traditional mathematics, "min" and "max" respectively (see [22, 78, 93]). This distinction between crisp sets and fuzzy sets becomes essential in the course of the decision-making process. Furthermore, there have been indications that in some situations of evidential reasoning the product operator may be preferred to the "min" operator even though we deal with fuzzy sets (see [22, 93]). Our goal is to explore the

utility of the possibility theory in evidential reasoning and to perform the decision-making process, which is basically defined over a fuzzified argument. Before doing this, we must, however, present some basic fuzzy mathematics as developed by Zadeh [84, 85, 86].

D.5 Averaging Operators

The average of two fuzzy sets is defined by mapping from $[0,1]^2 \rightarrow [0,1]$, such as $\forall x, y$, and $m \notin \{\min\ ,\ \max\}$,

1. $\min(x,y) \leq m(x,y) \leq \max(x,y)$,

2. $m(x,y) = m(y,x)$

3. m is a nondecreasing function of each of its arguments. From this definition, we state the following properties:

   (a) m is seen to be idempotent. Consequently, idempotence and the nondecreasing properties imply the inequality between m and either min or max.

   (b) If m is strictly increasing, then it cannot be associative.

   (c) The only associative means are the medians that are defined for a threshold value $\alpha \in [0,1]$ by the following [22]

$$\begin{aligned} \text{med}\ (x,y,\alpha) &= y \text{ if } x \leq y \leq \alpha \\ &= \alpha \text{ if } x \leq \alpha \leq y \\ &= x \text{ if } \alpha \leq x \leq y. \end{aligned}$$

   Therefore, the averaging operator does not imply associativity. Rather, it implies bi-symmetry, which is defined as

4. $\forall x, y, z, t,\ \ m[m(x,y), m(z,t)] = m[m(x,z), m(y,t)]$.

By these definitions $(1)-(4)$ and, if m is continuous and strictly increasing, the general form of the functional equations describing the mean has the following solution [2]:

$$m(x,y) \;=\; K^{-1}\left(\frac{K(x)+K(y)}{2}\right), \tag{D.35}$$

where K is continuous and strictly monotonic.

Symmetric Sums

Several aggregation operators for fuzzy sets were proposed by Silvert [73]. Those symmetric sums are self–dual in the DeMorgan sense. Consider a function, δ: $[0,1]^2 \rightarrow [0,1]$, which satisfies

1. $\delta(0,0) = 0$, $\delta(1,1) = 1$.

2. δ is commutative.

3. δ is a nondecreasing function of each argument.

4. δ is continuous.

5. $1 - \delta(x,y) = \delta(1-x, 1-y)$.

Every sum that satisfies conditions (1) through (4) and is symmetric in the sense of (5), can be put in the form [73]:

$$\delta(x,y) \;=\; \frac{g(x,y)}{g(x,y)+g(1-x,1-y)}, \tag{D.36}$$

where g is a continuous nondecreasing and nonnegative function verifying $g(0,0) = 0$. The following results can be easily obtained using Eq. D.36:

- $\delta(x, 1-x) = 1/2$.
- $\forall x$, if $g(0,x) = 0$, then $\delta(0,1) = 1/2$, where $g(0,x) = g(x,0)$.
- The only associative symmetric sum that is considered a mean lying between min and max is med $(x, y, 1/2)$.

Therefore, we cannot have associative and idempotent symmetric sums at the same time. That is, for associative symmetric sums, $\delta(0,1)$ is undefined. For instance, the associative sum that is generated by $g(x,y) = x \cdot y$ (*i.e.*, $g(0,x) = 0$) can be substituted in Eq. D.36 to yield

$$\delta_0(x,y) = \frac{xy}{1-x-y-2xy}. \tag{D.37}$$

An example of $g(x,y)$, which is not associative, is $g(x,y) = x+y-xy$ (*i.e.*, $g(0,x) \neq 0$). Substituting in Eq. D.36 yields

$$\delta_{\dot{+}}(x,y) = \frac{x+y-xy}{1+x+y-2xy}. \tag{D.38}$$

Let us now consider the idempotent symmetric sums, which, as we stated, are not associative. Since the idempotent symmetric sums are nondecreasing, they are averaging operators, *i.e.*, the mean of two fuzzy sets that lies between the min and the max, with $m(x,y) = m(y,x)$. If we include the property of bi-symmetry as it was defined previously and the property of strictly increasing, it can be verified that there is one single idempotent symmetric sum, namely, the arithmetic mean, generated by $g(x,y) = x+y$.

From Eq. D.36 we can choose g to express an intersection or union operation; *i.e.*, g being min or max yields

$$\delta_{\min}(x,y) = \frac{\min(x,y)}{1-|x-y|},$$

$$\delta_{\max}(x,y) = \frac{\max(x,y)}{1+|x-y|}. \tag{D.39}$$

These two operators are also means but not arithmetic means, which means that they are not bisymmetric. For $x+y \leq 1$, we have the following inequalities

$$\begin{aligned} \frac{xy}{1-x-y+2xy} &\leq \frac{\min(x,y)}{1-|x-y|} \leq \frac{x+y}{2}, \\ \frac{xy}{1-x-y+2xy} &\leq \frac{\max(x,y)}{1+|x-y|} \leq \frac{x+y-xy}{1+x+y-2xy}. \end{aligned} \tag{D.40}$$

For $x+y \geq 1$, the latter inequalities are reversed. These are some of the fuzzy set operations that are candidates for use in combining fuzzy assessments.

Combination Criteria from Fuzzy Premises

Let us assume that we have two sources of information, each of which gives an assessment represented by the possibility distribution, $\pi_i = \mu_{A_i}$, $i = 1,2$. One way to combine these two assessments is to compute the intersection of their possibility distribution and to normalize this result [22]:

$$\forall\theta \in \Theta,\ \pi_c(\theta) = \frac{\pi_1(\theta) * \pi_2(\theta)}{\sup\limits_{\theta\in\Theta} \pi_1(\theta) * \pi_2(\theta)}, \tag{D.41}$$

where $*$ is a fuzzy set–theoretic similar to the one we have already discussed, and $\pi_c(\theta)$ is the combined assessment. If $* = \min$, then Eq. D.41 becomes

$$\forall\theta \in \Theta,\ \pi_c(\theta) = \frac{\min[\pi_1(\theta), \pi_2(\theta)]}{\sup\limits_{\theta\in\Theta}[\min(\pi_1(\theta), \pi_2(\theta)]}.$$

By definition, the supremum is the least upper bound of a given set. Since we have $\pi(\bar{\theta}) = 1 - N(\theta)$, where $N(\theta)$ is the necessity measure, therefore, the combined possibility distribution can be written as

$$\forall\theta \in \Theta,\ \pi_c(\theta) = \frac{\min[\pi_1(\theta), \pi_2(\theta)]}{\max[\min(\pi_1(\theta), \pi_2(\theta)), \min[1 - N_1(\theta), 1 - N_2(\theta)]}.$$

Note that we assigned the value of the membership function $\mu_A(\theta)$ to be the possibility distribution, $\pi(\theta)$, which simplified the latter formula. Recall the complementation property as it was defined [84]:

$$\forall\theta \in \Theta \ , \ \mu_{\overline{A}}(\theta) = 1 - \mu_A(\theta) \, . \tag{D.42}$$

Recall also that

$$\pi(\theta) + \pi(\bar{\theta}) \geq 1 \, . \tag{D.43}$$

From Eqs. D.41 and D.42, we see that the case of $\pi(\theta) = \mu_A(\theta)$ is a special one of Eq. D.43; *i.e.*, $\pi(\theta) + \pi(\bar{\theta}) = 1$. Therefore,

$$\pi(\bar{\theta}) = 1 - N(\theta) = 1 - \pi(\theta) \, . \tag{D.44}$$

Substituting in the last expression of the combined possibility distribution, we have

$$\forall\theta \in \Theta, \ \ \pi_c(\theta) = \frac{\min[\pi_1(\theta), \pi_2(\theta)]}{\max[\min(\pi_1(\theta), \pi_2(\theta)), \min(1 - \pi_1(\theta), 1 - \pi_2(\theta))]} \, . \tag{D.45}$$

Another normalizing factor such as $1 - |\pi_1(\theta) - \pi_2(\theta)|$ was suggested by [73]. If the two sources of information are in severe conflict (conflicting evidence not in the sense of Dempster's rule of combination), then the $*$ operator can be selected as the union of the two proposed fuzzy sets. A similar procedure may be used to find the combined possibility distribution.

The results from applying the common example to the symmetric sum min operator are shown in Fig. D.13. In Fig. D.13(a), for the relative agreement case, the opinion is seen to lie between the two sources although it is closer to the larger source at high levels, and it is closer to the smaller source at low values. For the relative disagreement case in Fig. D.13(b), the operator results in a non-symmetrical output, which is near zero

whenever one of the sources is zero regardless of the other source value. The output is well below the maximum level of either source. When the sources are equal, the output is seen to be the same.

Figure D.14 shows the consensus for the symmetric sum max operator applied to the example. For the relative agreement case in Fig D.14(a), again as in the min operator, the opinion lies between the two sources although in this method the output is nearer the smaller source at high levels and nearer the larger source at low levels. Considering the relative disagreement Fig. D.14(b) shows considerably different results from the min operator. The output is seen to be symmetrical, tending to follow the larger of the two source values, although the output is well below the maximum value of either source. Again, when the sources are equal, the output is the same.

Figure D.15 gives the output for combination criteria from fuzzy premises. In the relative disagreement case, the output is seen to be one for larger values of the sources while it follows the lower of the two sources at smaller levels. Figure D.15(b) shows the output for the relative disagreement case. The result is a non-symmetric opinion that lies well below the maximum value of either source. Zero output is approached when one of the sources is zero.

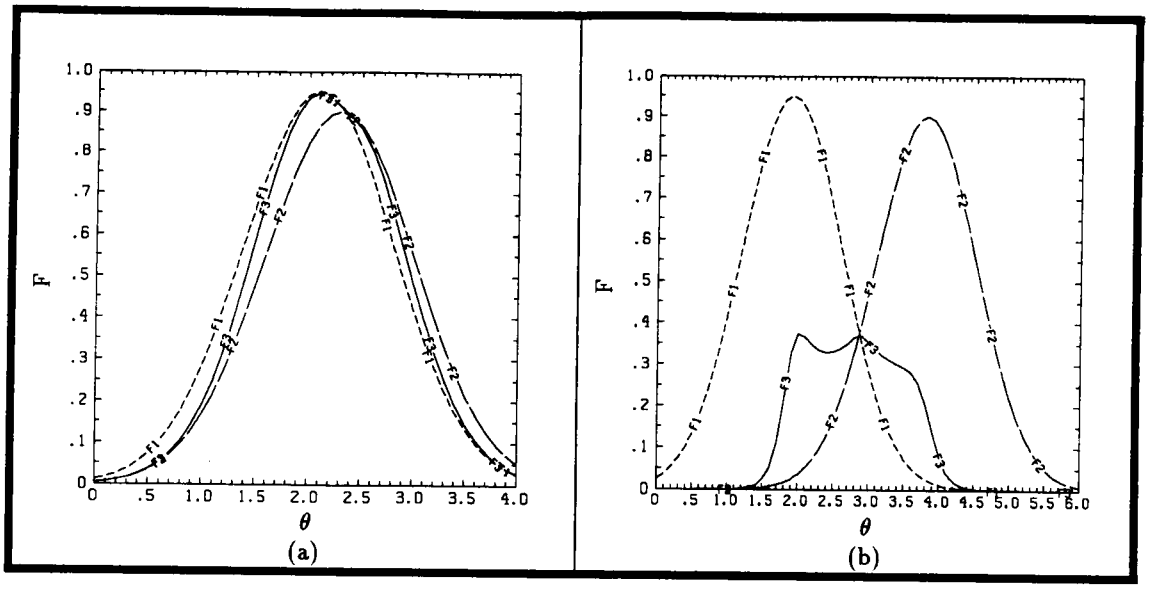


Figure D.13: *(a) the symmetric sum: min operator experiment $\mathcal{A}$; (b) the symmetric sum: min operator for experiment $\mathcal{B}$.*

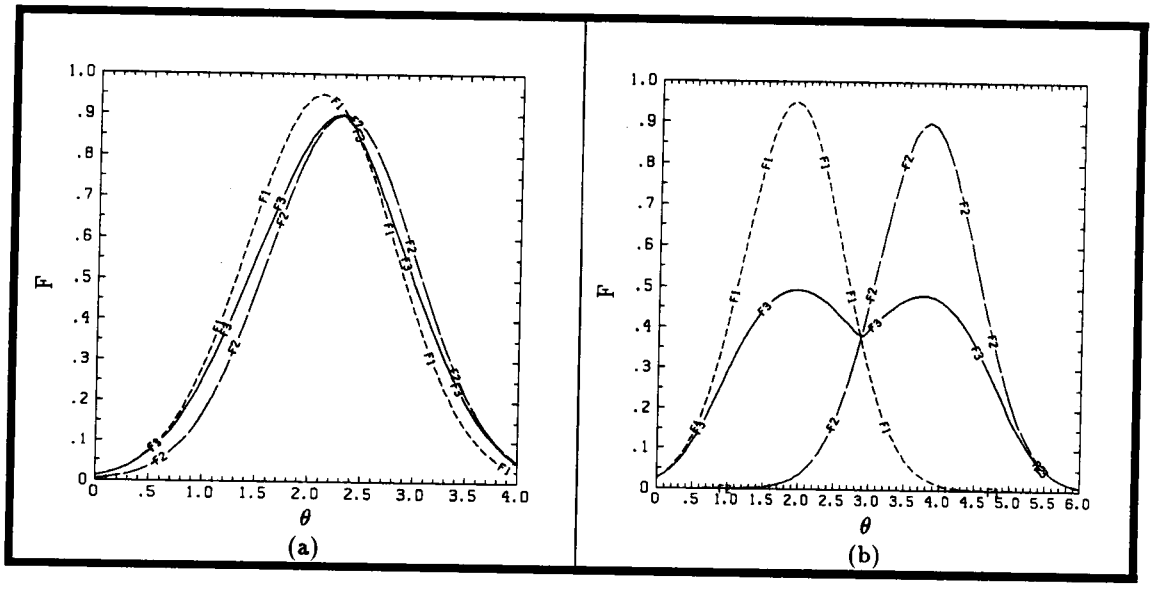


Figure D.14: *(a) the symmetric sum: max operator experiment $\mathcal{A}$; (b) the symmetric sum: max operator for experiment $\mathcal{B}$.*

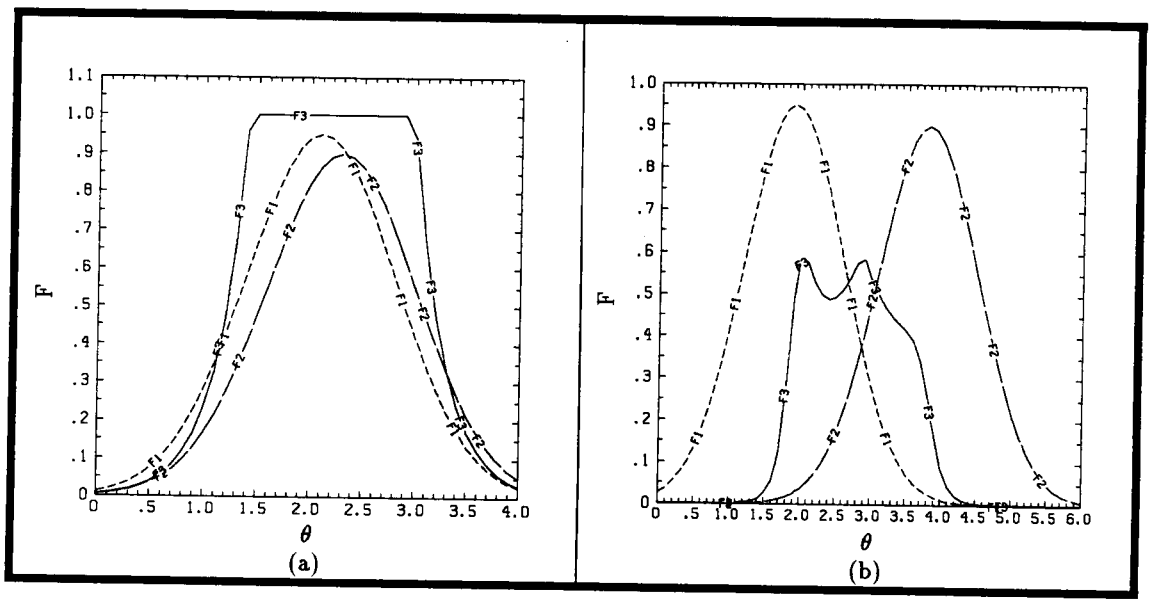


Figure D.15: *(a) combination criteria from fuzzy premises for experiment A; (b) combination criteria from fuzzy premises for experiment B.*

APPENDIX E

Comments on Appendices A – D

In appendices A – D, we have reviewed and compared the principal data fusion methods which use uncertainty and imprecision. Probability theory, evidence theory, and possibility theory have each been examined in detail. A common example has been applied to each method in order to provide a direct comparison of the different approaches. From the example, one test case had two opinions which were in agreement while the other case had two opinions in disagreement. These two extremes of data were used to illustrate the performance ranges for each method.

We have discussed some limitations of the probability theory and particularly the Bayesian approaches to solve dispute problems between several sources of information. Briefly, *a priori* knowledge, which is assumed in most cases, introduces unfair treatment among bodies of evidence to be combined. Representing ignorance introduces inconsistency of assumptions and, consequently, a misleading inference might occur. Conditioning in favor of the *a priori* knowledge does not perfectly fit the criteria of combining evidence. The approaches that commit themselves to obey the rules of the probability theory are mentioned, and the performances of such approaches are shown. All methods such as the L.O.P, I.O.P, H.O.P, and the Super Bayesian Approach work somewhat satisfactorily in the case where information sources are in relative agreement. In contrast, the problem becomes more complex when the information sources have clear dispute. All the methods mentioned above do not have tools to find which source

of information is more reliable than another. Rather, these methods actually perform an averaging process to a certain extent. The process is usually governed by the weight of each assessment. In many cases, partial reinforcement is desirable so that one must avoid the use of the L.O.P. since there is no reinforcement of opinion. This does not mean that the I.O.P. or the H.O.P. is preferrable. If a partial reinforcement of opinion is desirable, the independent opinion pool can be too extreme in this concern [7]. The only method, however, that has a different way of inference is the Super Bayesian Approach, which conditions the likelihood (proposition) using the *a priori* knowledge (decision maker). The latter is also unable to generate satisfactory inference between pieces of information that possess a wide margin of dispute. Although Dempster's rule of combination seems to solve conflicting bodies of evidence, there are still some weaknesses in doing so in many situations. As we have shown in Figs. D.3 through D.6 and in Fig. D.12, all methods perform relatively well where the proposed bodies of evidence are in relative agreement. In contrast, all of them including Dempster's rule of combination, do not perform well in the case of sharp conflicting bodies of evidence.

Unlike Bayesian theory, the theory of evidence (belief) does not allow *a priori* degree of belief inevitably to dominate any statistical inference. The theory of belief permits nonadditive degrees of belief and emphasizes the combination of evidence instead of conditioning. It reveals consistency in representing ignorance while the Bayesian theory suffers from this dilemma. From a statistical point of view, the theory of belief is not problem-free either. For instance, if a support function, S_1, is committed to a proposition, A, and a support function, S_2, is committed to the same proposition, A, then the orthogonal sum, $S_1 \oplus S_2$, is a new belief function committed to A. This orthogonal sum, $S_1 \oplus S_2$, that we have derived in Eq. D.25 is always greater than either support func-

tions committed to A. More specifically, if $S_1 = 0.9$ and $S_2 = 0.05$, then $S_1 \oplus S_2 > S_1$. This tells us that even though S_2 is a very weak support function, it will increase the combined belief committed to A. This makes the theory of belief give an overestimation of judgmental reasoning that is undesirable in most cases. Many problems can arise in using Dempster's rule of combination. The combined evidence strength is very sensitive to the numerical values of the inputs; *i.e.*, a small change of the input value could cause a drastic change of the output value as pointed out by Zadeh [91] and illustrated in the following example.

Example [90]: Let a, b, and c be mutually exclusive events (*i.e.*, a, b, and c cannot happen simultaneously). Assume that we have two information sources, m_1 and m_2:

$$m_1(a) = 0.01 \ , \ m_1(b) = 0.1 \ , \ m_1(c) = 0.89$$

$$m_2(a) = 0.89 \ , \ m_2(b) = 0.1 \ , \ m_2(c) = 0.01.$$

Using Dempster's rule of combination given in Eq. D.24, we get

$$m(a) = 0.32 \ , \ m(b) = 0.36 \ , \ m(c) = 0.32.$$

Now, let us assume a slight change of the input data:

$$m_1(a) = 0 \ , \ m_1(b) = 0.1 \ , \ m_1(c) = 0.9$$

$$m_2(a) = 0.9 \ , \ m_2(b) = 0.1 \ , \ m_2(c) = 0.0.$$

Using Dempster's rule of combination in Eq. D.24, we get

$$m(a) = 0.0 \ , \ m(b) = 1.0 \ , \ m(c) = 0.0.$$

Notice that Dempster's rule of combination shows extreme sensitivity to a slight change of the input.

Regarding Figs. D.13 through D.15. It is clear that an aggregation using fuzzy sets is different from what we have traditionally seen using crisp sets. We have seen the performance of the probability theory and the theory of belief, then a comparison analogy has been made. Although the symmetric sum is a fuzzy concept, which employs both min and max operators, it shares the same dilemmas of an averaging process that we have already shown using the linear opinion pool. Hence, if we are given contradicting bodies of evidence, the fused data are not so encouraging. In contrast, Fig. D.15 shows an aggregation, which computes the intersection of the possibility measures. In this case, a normalization factor, which involves the *supremum* operator, is used. In the event that there is a sharp dispute between the given data, this technique shows some reconciliation role in a different way; *i.e.*, it manages to involve both the *credibility* and *possibility* measures. Hence, the outcome not only reflects the strength of an assessment, but also its confidence. However, we must mention that the combined evidence is overestimated when the proposed bodies of evidence are both very fuzzy.

VITA

The author of this dissertation, Muhamad B. Abdulghafour, was born in Elhassekeh, Syrian Arab Republic on May 20, 1955. He graduated from Abo–Ther El–Ghefary High School, Elhassekeh, Syria, in 1972. He received his Bachelor of Science degree in electronics engineering from Menofia University, Menouf, Egypt, in May 1977. After graduation, he was employed by the Syrian Petroleum Company for six years. In spring of 1984, he came to the University of Tennessee, Knoxville (UTK), to work on his Master of Science degree. After the completion of the requirements for the Master of Science in summer 1986, he continued his study toward the degree of Doctor of Philosophy, with a major in electrical and computer engineering. While working toward this degree in UTK, he served as a graduate research assistant in the areas of computer vision, pattern recognition, robotics, and artificial intelligence in the Electrical and Computer Engineering Department. He has been an author and a co–author of several papers and technical reports in his areas of research.

The author is a member of several technical associations such as IEEE, Eta Kapa Nu, SPIE, and the Association of Syrian Engineers.