Distributed Detection and Fusion in Parallel Sensor Architectures

A Thesis

Submitted to the Faculty

of

Drexel University

by

Sayandeep Acharya

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

November 2014

Dedications

"If I cease searching, then, woe is me, I am lost. That is how I look at it - keep going, keep going come what may."

– Vincent van Gogh, *The Letters of Vincent van Gogh*, July 1880

To Ma, Baba, Didi and Diya...

Acknowledgments

This dissertation is the culmination of the most enlightening journey of my life. Throughout this wonderful adventure, there were some individuals who were my continuous source of motivation and support. It is humbling to reflect on their sacrifices for me and I would like to express my sincere gratitude and respect for them.

I sincerely thank my advisor Dr. Moshe Kam for his constant guidance and encouragement throughout my years of study at Drexel University. I am and will forever remain in awe of his knowledge, humility and infallible dedication to his craft and would always cherish everything I learned while working with him. He had always been the friend, philosopher, guide I needed to get through graduate school.

I cannot over emphasize the role my parents and my sister played in pushing me to be who I am today. I owe to them a debt of gratitude for their love, support, understanding and for being the constant source of motivation.

I am thankful to my colleagues and friends in Data Fusion Laboratory for the many memorable moments and stimulating discussions. Finally, I would like to thank the other members of my dissertation committee, Dr. Paul Kalata, Dr. Leonid Hrebien, Dr. Matthew Stamm, Dr. Chris Rorres for their advice and support.

Table of Contents

List of Tables

List of Figures

List of Symbols

Symbol	Reference	Description
$\|A\|$		Cardinality of set A
$\overline{A}$		Complement of set A
$A \cup B$		Set union of sets A and B
$A \cap B$		Set intersection of sets A and B
H_i		i^{th} Hypothesis
$P(A)$		Probability of event A
$E(.)$		Expected value
$P(H_i)$		Prior probability of hypothesis H_i
N		Number of sensor/detectors in a distributed fusion system
z_i		Raw observation of the i^{th} sensor/detector
$\psi(z_i)$	Section 2.1	Local detector decision rules
u_i	Chapters 2, 3, 8	Binary local decision of i^{th} detector
u_0	Chapters 2, 3, 8	Binary global decision of decision fusion center
P_{F_i}	Chapters 2, 3, 8	False alarm rate of i^{th} local detector
P_{D_i}	Chapters 2, 3, 8	Detection rate of i^{th} local detector
P_{F_0}	Chapters 2, 3, 8	False alarm alarm rate of decision fusion center
P_{D_0}	Chapters 2, 3, 8	Detection rate of decision fusion center
α	Section 2.3, Chapter 3	Chosen global false alarm probability under Neyman-Pearson criterion

t_g	Sections 2.3, 3.1	Global likelihood ratio test threshold
t_{loc}	Chapter 3	Local detector likelihood ratio test threshold
$\Lambda(u)$	Section 3.1	Global likelihood ratio at fusion center as a function of local decisions
$P(X\|H_i)$	Section 3.1	Conditional distribution of random variable X under hypothesis H_i
p	Chapter 3	Local false alarm rate for identical detectors
q	Chapter 3	Local mis-detection rate for identical detectors
γ	Chapter 3	Convex combination coefficient for randomized Neyman-Pearson test
Θ	Chapter 4	Frame of discernment under Dempster-Shafer theory
θ_i	Chapter 4	A particular element of the frame of discernment
ϕ	Chapters 4, 5, 6	Null element in the power set of the frame of discernment
$b(X)$	Chapters 4, 5, 6	Belief function toward proposition X
$d(X)$	Chapter 5	Disbelief toward proposition X
$u(X)$	Chapter 5	Evidential uncertainty toward proposition X
$pl(X)$	Chapters 4, 5, 6	Plausibility toward proposition X
r_i	Chapter 4	Reliability of the i^{th} soft source
$BetP(X)$	Chapters 4, 5, 6	Pignistic probability transformation
$BetP^{\Delta}(X)$	Chapter 6	Non-normalized Pignistic probability transformation over a frame Δ
$\oplus$	Chapter 4	Dempster's belief combination operator
$\odot$	Chapters 4, 6	Conjunctive belief combination operator

$\otimes$	Chapter 4	Disjunctive belief combination operator
ω_X	Chapter 5	Opinion tuple toward proposition X
S_i	Chapters 5, 6	i^{th} soft source
$\omega_X^{S_i}$	Chapter 5	Opinion tuple toward proposition X held by soft source S_i
$a(X)$	Chapter 5	Relative atomicity of proposition X
$a(X\|Y)$	Chapter 5	Conditional relative atomicity of proposition X given proposition Y is true
$Pe(X)$	Chapter 5	Probability Expectation
$\widetilde{\Theta^X}$	Chapter 5	Focused frame of discernment with focus on proposition X
g_i	Chapter 5	Posterior detection probability of hypothesis H_1 from i^{th} local detector
(ν, ρ)	Chapter 5	Symmetric Beta distribution parameters to model soft and hard sensor support assignments
Ω	Chapter 6	Root frame in evidence tree
$\Theta^i[P]$	Chapter 6	Child frame at level $i > 0$ under parent frame P
$m_{\Theta^j[P]}(X\|P)$	Chapter 6	Scaled belief mass assigned toward subset X of child frame $\Theta^j[P]$
$m_{\Delta h}(X)$	Chapter 6	Bayesian belief mass assigned toward subset X of a frame Δ by a hard source
$m_{\Delta s}(X)$	Chapter 6	Bayesian belief mass assigned toward subset X of a frame Δ by a soft source
Ψ	Chapter 6	Consolidated frame or decision layer at some level of the evidence tree
Pr	Chapter 6	Pignistic probability vector on a decision layer
$w(k)$	Chapter 9	Process noise vector at time k

$v(k)$	Chapter 9	Measurement noise vector at time k
$G(k)$	Chapter 9	Process noise matrix at time step k in state dynamics equation
$B(k)$	Chapter 9	Kalman filter input matrix at time step k
$H(k)$	Chapter 9	Kalman filter observation matrix at time step k
$Q(k)$	Chapter 6	Process noise covariance matrix at time k
$R(k)$	Chapter 6	Measurement noise covariance matrix at time k
$(k\|k-1)$	Chapter 9	Denotes quantity at time k given a measurement at time $k-1$
$(k\|k)$	Chapter 9	Denotes quantity at time k given a measurement at time k
$P(k\|k)$	Chapter 9	State estimation error covariance matrix at time k given a measurement at time k
$P(k\|k-1)$	Chapter 9	State estimation error covariance matrix at time k given a measurement at time $k-1$
a_1,a_2	Chapter 9	Second order auto-regressive model parameters for temporally correlated measurement noise
$e(k)$	Chapter 9	Zero mean Gaussian noise at time k driving the AR model of measurement noise.

Abstract

Distributed Detection and Fusion in Parallel Sensor Architectures
Sayandeep Acharya
Moshe Kam, Ph.D.

Parallel distributed detection system consists of several separate sensor-detector nodes (separated spatially or by their principles of operation), each with some processing capabilities. These local sensor-detectors send some information on an observed phenomenon to a centrally located Data Fusion Center for aggregation and decision making. Often, the local sensors use electro-mechanical, optical or RF modalities and are known as "hard" sensors. For such data sources, the sensor observations have structure and often some tractable statistical distributions which help in weighing their contribution to an integrated global decision. In a distributed detection environment, we often also have "humans in the loop.". Humans provide their subjective opinions on these phenomena. These opinions are labeled "soft" data. It is of interest to integrate "soft" decisions, mostly assessments provided by humans, with data from the "hard" sensors, in order to improve global decision reliability. Several techniques were developed to combine data from traditional hard sensors, and a body of work was also created about integration of "soft" data. However relatively little work was done on combining hard and soft data and decisions in an integrated environment.

Our work investigates both "hard" and "hard/soft" fusion schemes, and proposes data integration architectures to facilitate heterogeneous sensor data fusion. In the

context of "hard" fusion, one of the contributions of this thesis is an algorithm that provides a globally optimum solution for local detector (hard sensor) design that satisfies a Neyman-Pearson criterion (maximal probability of detection under a fixed upper bound on the global false alarm rate) at the fusion center. Furthermore, the thesis also delves into application of distributed detection techniques in both parallel and sequential frameworks. Specifically, we apply parallel detection and fusion schemes to the problem of real time computer user authentication and sequential Kalman filtering for real time hypoxia detection.

In the context of "hard/soft" fusion, we propose a new Dempster-Shafer evidence theory based approach to facilitate heterogeneous sensor data fusion. Application of the framework to a number of simulated example scenarios showcases the wide range of applicability of the developed approach. We also propose and develop a hierarchical evidence tree based architecture for representing nested human opinions. The proposed framework is versatile enough to deal with both hard and soft source data using the evidence theory framework, it can handle uncertainty as well as data aggregation.

1: INTRODUCTION

Data Fusion is the process of combining information from several different sources pertaining to the same event, environment or phenomenon. The objective is to develop a robust and more complete description of the environment or process of interest than would be normally available with data from a single source. The field of data fusion is of significance in any application where a large amount of data must be combined, fused and distilled to obtain information of appropriate quality and integrity on which decisions can be made. Data fusion finds application in many military systems, in civilian surveillance and monitoring tasks, in process control and in information systems.

1.1 Fusion Objectives and Sensor Types

The basic intuition behind incorporating multiple information sources to collect information is that the aggregated data might be more reliable (less noisy) and therefore can aid in better understanding of the phenomenon under surveillance. Typically, the fusion objectives of a specific application scenario include one or more of the following functions [1]:

- Detecting presence of an object or environmental condition
- Object recognition and classification
- Target tracking

- Health monitoring and flagging changes
- Intelligent decision making and situation assessment

If sensors used to collect observations merely duplicate information acquisition, the fusion process essentially incorporates redundancy for enhancing reliability. This situation might not facilitate better understanding of the phenomenon in question. Therefore, most multi-sensor fusion systems incorporate heterogeneous sensors so that a wide range of information with varying degrees of uncertainty can be collected and fused for end decision making. Under the broader perspective of considering sensors as information sources, such heterogeneous multi-sensor systems can have humans as sources as well.

According to current terminology [2], sensors which are traditional in the mode of operation and have well defined statistical error characteristics like electro-mechanical or electro-optic sensors are termed as *Hard* sensors where as human sources which typically produce ambiguous or imprecise information are termed as *Soft* sensors. The classical field of fusion deals with hard sensors and the task becomes that of developing probabilistic algorithms for data fusion. On the other hand, since human opinions are subjective and very difficult to mathematically model, probabilistic approaches generally fail to reliably define fusion systems with humans as information sources. As is presented in Chapter 4, in such situations, tools from evidence theory can be used to facilitate fusion of Hard and Soft sensors.

1.2 Data Fusion Models

In any data fusion problem, there is an environment, process or quantity whose true value, situation or state is unknown. It would be unreasonable to expect that there is some single source of perfect and complete knowledge about the problem of interest and so information must be obtained indirectly from sources which provide imperfect and incomplete knowledge, using these to infer the information needed. In the fusion problems we study, the state of the environment is represented as a set of hypothesis denoted by H. The binary hypothesis detection problem deals with the situation when the environment is assumed to be either in *state 0* denoted by hypothesis H_0 or *state 1* denoted by hypothesis H_1. Since most multi-hypothesis situations can be represented in terms of a hierarchy of binary hypotheses, all the algorithms described herewith assumes the environment to be modeled by a binary hypothesis set.

The data fusion systems studied here involve multiple sensors or detectors collecting observations (henceforth called local observations) about the environment, and transmitting a processed version of the observation to a fusion center which is responsible for data aggregation. The end goal is to combine the received data in an optimal or near-optimal form so that a reliable and informed decision can be made about which hypothesis might be true. In a binary hypothesis detection problem, the decision on the set of hypotheses is represented as follows

$$u = \begin{cases} 1, & \text{if } H_1 \text{ is accepted} \\ -1, & \text{if } H_0 \text{ is accepted.} \end{cases} \tag{1.1}$$

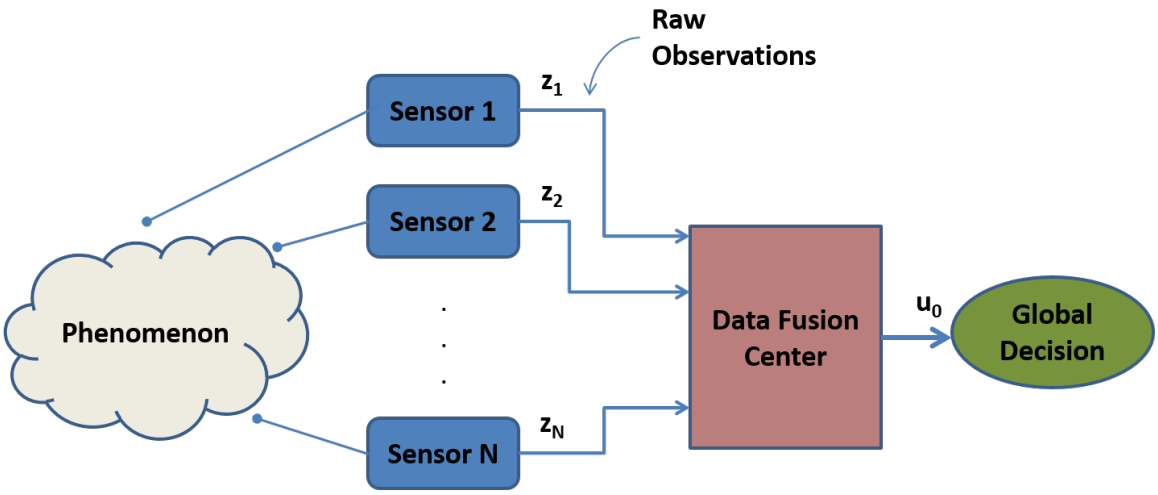


Figure 1.1: A centralized data fusion system.

1.3 Data Fusion Architectures

We consider a set of sensors physically distributed around an environment. In a *centralized* data fusion system, the raw sensor observations are communicated to a central fusion center that solves a classical hypothesis testing problem and decides on one the possible hypotheses. Figure 1.1 shows the basic scheme for a centralized fusion system.

In this scheme, the N local sensors do not perform any significant data compression; they send forward their raw observations z_i, $i = 1...N$ to the central fusion center which then combines the incoming data to produce a global decision. A distinct alternative, (e.g., [3]), is a *distributed* [1] data fusion system, where each sensor has an associated local processor which can extract useful information from the raw sensor observations prior to communication. A summary of the local observations

[1] Often the configuration introduced in [3] and its variations are inappropriately also referred to as decentralized, however there are some significant architectural differences between decentralized and distributed configurations [4].

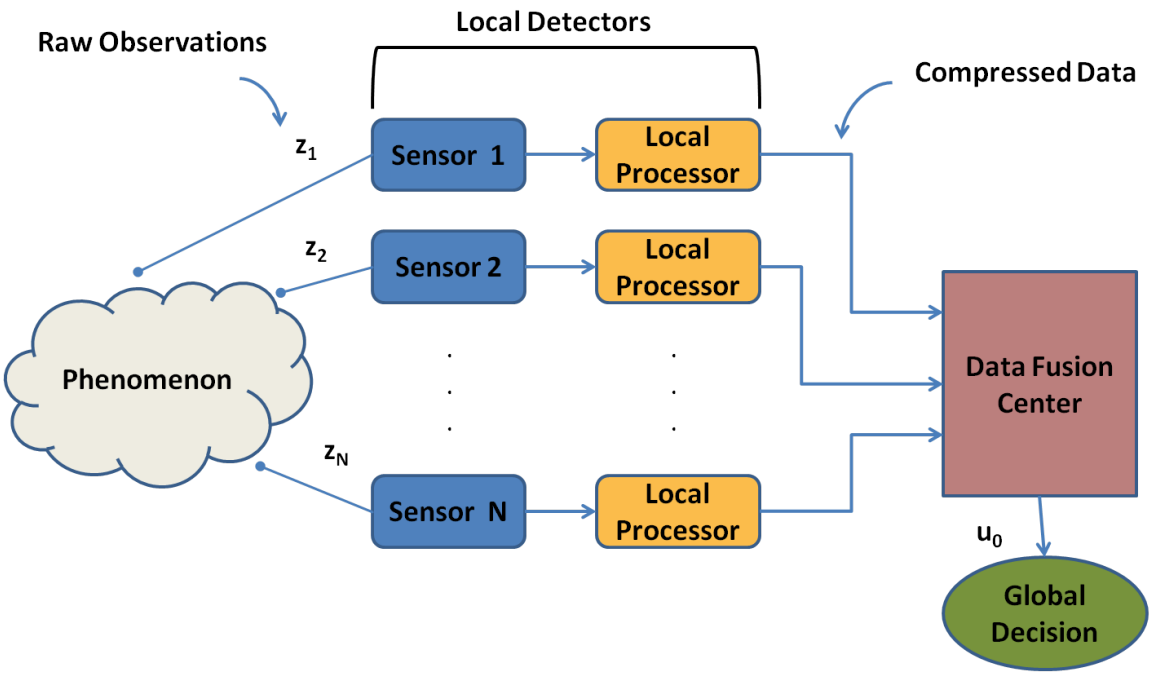


Figure 1.2: A distributed data fusion system.

(test statistics) is sent to the fusion center which then makes a decision on the basis of the messages received. The fusion center faces an hypothesis testing problem (the messages received from the local sensors are considered as the fusion center's observations). Additional issues that need to be addressed are the signal processing schemes at the individual local sensors and the nature of information transmitted from the sensors to the fusion center.

The move to more distributed, autonomous, organizations is clear in many information processing systems. This is most often motivated by two main considerations; the desire to make the system more modular and flexible, and a recognition that a centralized structure imposes unacceptable overheads on communication and central computation. Figure 1.2 shows the basic scheme for a distributed fusion system.

In a distributed fusion scheme, the intermediate local processors may make hard decisions and transmit these results to the fusion center for decision combination. Other options include transmission to the fusion center of multi-level decisions instead of binary decisions, transmission of only changes from a baseline or previous decision, transmission of quantized observations. In all of these cases, the intermediate data compression leads to information loss. Hence even though the distributed scheme is modular, easier to implement and has much less communication bandwidth requirements, it almost always has suboptimal performance (irrespective of the performance metric) as compared to a centralized architecture where the fusion center works with all available information. For any fusion system design, the centralized scheme can therefore be assumed to provide an upper bound on the performance and serve the standard for comparative analysis. We are facing here a tradeoff between detection performance and the required information storage, communication and processing required to achieve this performance.

Several different topologies all of which fall under the umbrella of distributed fusion have been proposed. They include the following.

- **Parallel Decision Fusion**: In this scheme, shown in Figure 1.3, the local sensors form a bank of data collection nodes which map their observation vectors to local decisions u_i, $i = 1...N$. These are then sent forward through dedicated communication channels to the decision fusion center which then processes the received local decisions and produces a global decision on the set of hypotheses. For a binary hypothesis set, the local and global decisions can be represented

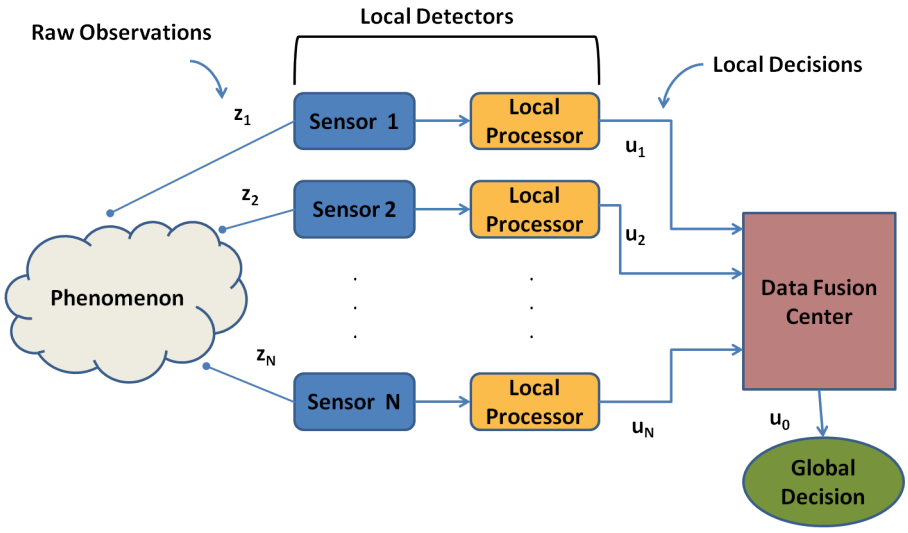


Figure 1.3: A parallel decision fusion system.

with 1 bit and are of the form as shown in (1.1). This architecture is by far the most studied distributed fusion topology and all the results presented in this thesis assume a similar parallel distributed decision fusion system.

Variations of the topology shown in Figure 1.3 sometimes incorporate additional features like feedback from fusion center ([5–7]) as well as inter sensor communication [8]. An excellent survey of the most seminal works in the field of distributed detection can be found at ([1], Chapter 5).

- **Sequential or Tandem Fusion**: In serial configuration of N sensors, the $(j-1)^{th}$ sensor passes its quantized information to the j^{th} sensor which generates its quantized information based on its own observation and the quantized data received from the *previous* sensor in the sequence [9]. The first sensor in the network uses only its own observations to compute its quantized data for use by the next sensor. The last sensor in the network acts as the fusion center and

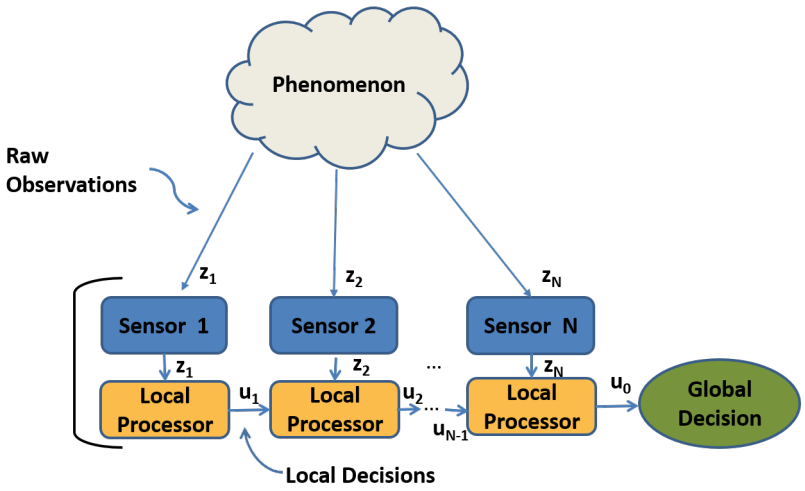


Figure 1.4: A sequential data fusion system.

makes the final decision on the set of hypotheses. Figure 1.4 shows a sequential fusion topology.

A disadvantage of the serial topology is the delays that may accumulate, since each sensor has to wait for a decision from the previous sensors. Furthermore, there is the inherent problem of managing sensor failure, as a malfunctioning sensor anywhere in the sequence would affect functioning of the entire system. Another issue is the ordering of nonidentical sensors in serial networks. It might seem logical to put the best detectors toward the end but counter examples show that this configuration need not always be optimal [10].

- **Tree Structure**: Several tree topologies with hierarchical structures were proposed. The sensors at the lowest level of the tree send their processed information to the parent sensors who use their own observations and the information received from child sensors to compute their own summarized data which then

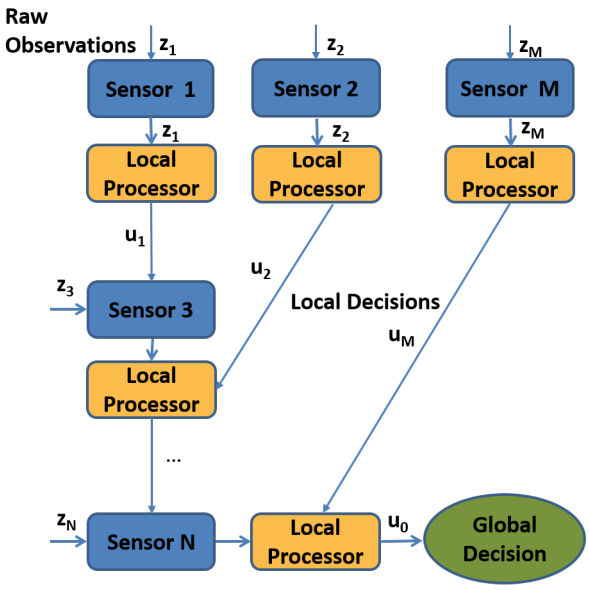


Figure 1.5: A Tree topology for distributed detection.

move up the hierarchy to the sensor at the root location which generates the global decision. Figure 1.5 shows an example tree topology.

There are many other hybrid structures that have been developed over the years each having their own assumptions and contexts. Results on distributed detection using some of the above topologies are discussed in ([11], Chapters 3, 4, 6).

1.4 Multi-Sensor Estimation

Central to the problem of Data Fusion as described in the previous sections, is the issue of estimation. Fundamentally, an estimator is a decision rule which takes as an argument a sequence of observations and whose action is to compute a value for the parameter or state of interest. Almost all data fusion problems involve this estimation process: we obtain a number of observations from a group of sensors and using this information we wish to find some estimate of the true state of the environment we

are observing. This estimate may be in the form of a binary decision toward some hypothesis (as discussed in Section 1.2) or a real number signifying the value of some chosen parameter. Estimation encompasses all important aspects of the data fusion problem. Sensor models are required to understand what information is provided, environment models are required to relate observations made to the parameters and states to be estimated, and some concept of information value is needed to judge the performance of the estimator (reduction of uncertainty in estimate). In other words, the estimation problem is central in linking the real world as observed by a sensor to the decisions we make about how to control or influence our environment.

Instead of explicit decisions toward hypotheses, when the goal is to compute numeric estimates of certain quantities (e.g., physical attributes like position, speed, percentage of concentration etc.) from noisy observations, one of the most versatile estimation and fusion algorithms is the Kalman filter [12]. The Kalman filter is a recursive linear estimator which successively calculates an estimate for a continuous valued state, that evolves over time, on the basis of periodic observations that of this state. The Kalman filter employs an explicit statistical model of how the parameter of interest evolves over time and an explicit statistical model of how the observations that are made are related to this parameter. A detailed coverage of existing techniques based on the Kalman filter is provided in [13–15]. In Chapter 9 of this thesis, we pose the problem of Hypoxia detection under the context of distributed detection and showcase the use of Kalman filter for data fusion.

1.5 Thesis Overview

This thesis is organized in three parts.

Part I deals with *Hard* fusion. The local sensors possess statistical error models and have known ranges of operation. **Chapter 1** introduces the field of Data Fusion, discusses the various architectures and topologies used in practical system design, and describes the classical distributed decision fusion problem. **Chapter 2** discusses the development and evolution of the distributed (parallel) decision fusion problem and the various probabilistic frameworks that are used to solve the same. Previous work in the area is presented. **Chapter 3** presents an original contribution of the thesis, an algorithm to compute the optimal distributed decision fusion system operating point under the Neyman-Pearson criterion.

Part II continues the study of parallel decision fusion architectures as it derives useful when humans also act as information sources. In such situations, the local sensors are called *Soft* sensors. We identify algorithms that fuse both *Hard* and *Soft* sensors data (*Hard/Soft* fusion). **Chapter 4** introduces the field of Evidence Theory (sometimes also referred to as Dempster-Shafer Theory [16]) and defines the *Hard/Soft* fusion problem. In **Chapter 5**, we present a new evidence theory based architecture that can facilitate Hard/Soft fusion. The proposed algorithm is discussed and evaluated against several scenarios. **Chapter 6** presents a new hierarchical evidence tree structure that can be used to mathematically model nested, equivocal or tentative human assessments. **Chapter 7** summarizes the thesis contribution on *Hard/Soft* fusion and offers future avenues of research.

Part III introduces the application of two classical and well established detection and fusion algorithms on couple of real world problems (algorithms implemented on real data sets).

Specifically, **Chapter 8**, discusses the application of parallel decision fusion techniques to perform identification of an user of a computer using biometric sensors. The problem we attempt to address is real time identification of whether an user is legitimate or unauthorized, using information from behavioral biometric systems (e.g., keyboard and mouse usage). This work poses the problem under the paradigm of distributed detection and also identifies the extent of marginal contributions of groups of biometric sensors toward the end fused result, whether an user is authentic or not.

Chapter 9 illustrates a Kalman filter based parallel fusion system that can aid in reliable and real time hypoxia (diminished availability of oxygen) detection in individuals exposed to high altitude environments. The fusion scheme is tested on actual data collected from multiple test subjects.

Chapter 10 presents the concluding remarks and summarizes the main contributions of the thesis.

Part I

DISTRIBUTED DETECTION - CLASSICAL FRAMEWORK

2: CLASSICAL DISTRIBUTED DETECTION AND FUSION

In many practical situations, one is faced with a decision making problem of choosing a course of action from a set of possible alternatives. Such situations arise in several applications ranging from radar detection, medical diagnosis, pattern recognition to stock market predictions. The general idea is to arrive at a choice represented by one among a set of hypotheses $H_1, ..., H_M$, based on information obtained from data which are invariably corrupted by noise and therefore uncertain.

2.1 Distributed Detection without Fusion

Distributed detection with parallel topology is the direct extension of hypothesis testing to the problem where instead of a single decision maker, there are N decision makers all of which decide individually on a particular hypothesis. For the remainder

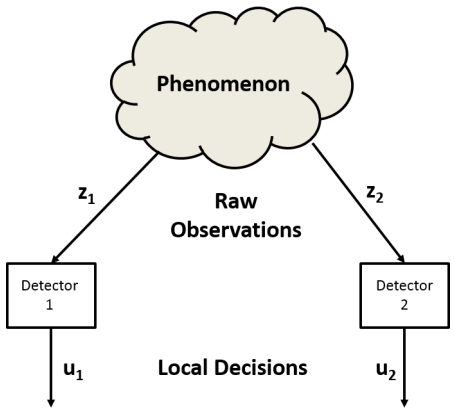


Figure 2.1: Parallel distributed detection without fusion.

of this study, we will focus only on binary hypothesis testing with the two hypotheses denoted as H_0 and H_1. The inception of the field of distributed detection as we are familiar today can be largely attributed to the seminal work of Tenney and Sandell [3]. The architecture used in [3] is shown in Figure 2.1 where two local sensors/detectors observe a phenomenon and collect noisy observations z_i, $i = 1, 2$.

The objective is to obtain rules $\psi_i(z_i)$ at each detector which are functions of local observations and produce local binary decisions u_i as shown in (1.1) such that the average cost of decision making (Bayes' risk) is minimized. The main result in [3] showed that the local decision rules are not standard likelihood ratio tests but comprise of coupled nonlinear equations (threshold of detector 1 is dependent on threshold of detector 2 and vice-versa) with data dependent thresholds which must be solved simultaneously to obtain the local decision regions. As shown in [17], this problem of distributed detection is NP complete. The situation gets better under the assumption that the local observations z_i are statistically independent conditioned on the hypotheses. In that case the decision rules though still coupled no longer have data dependent thresholds and can be numerically solved. These coupled equations are the Person-By-Person optimality conditions for the distributed detection problem. Tsitsiklis and Athans in [17] reported a study on the computational complexity of discrete models of some basic distributed detection/decision problems. Later the work in [18] extended the results in [3] for multiple hypothesis and multiple sensors.

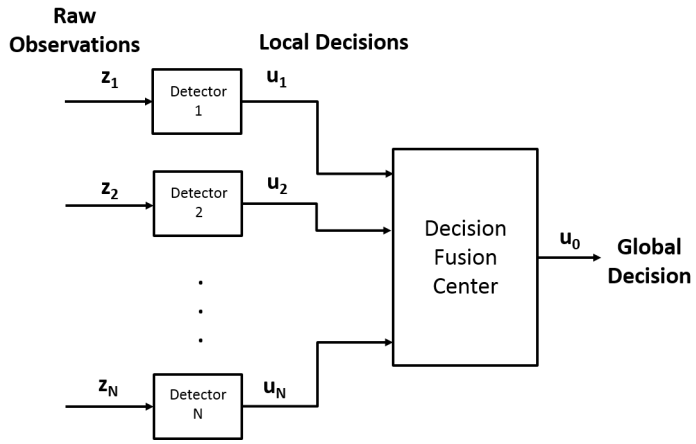


Figure 2.2: Parallel distributed detection with fusion.

2.2 Distributed Detection with Fusion

In the distributed detection system shown in Figure 2.1, the local decisions were not sent to a fusion center for global decision making. The addition of fusion center adds to the complexity of the system but guarantees better reliability in the certainty of the end decision. A parallel distributed decision fusion system with N local detectors is shown in Figure 2.2.

The local decision rules $\psi_i(z_i)$ map the local observations z_i to binary decisions u_i, $i = 1, ..., N$ which are then combined by the fusion center to produce a global decision u_0 such that

$$u_0 = \begin{cases} 1, & \text{if } H_1 \text{ is accepted globally} \\ -1, & \text{if } H_0 \text{ is accepted globally.} \end{cases} \tag{2.1}$$

In this classical set up, two inherently different problems need to be considered: the design of the fusion center decision rule, which strives for an optimal system

performance using compressed inputs from distributed sensors, and the design of local sensor decision rules. These two problems are intertwined with each other and need to be jointly handled to optimize a specified performance criterion.

2.3 Optimization Criterion

The design of any distributed detection system with or without fusion requires the solution of decision rules which are optimal in the sense of minimizing some chosen optimization criterion. When the observations collected at all local detectors are assumed to have defined probability distributions, the design of a distributed detection system may follow a probabilistic framework (as opposed to fuzzy or evidence theoretic frameworks). The performance of the fusion center and the local detectors are measured using the *probability of false alarm* (declare H_1 when H_0 is true): $P_{F_i} = P(u_i = 1|H_0)$ and *probability of detection* (declare H_1 when H_1 is true): $P_{D_i} = P(u_i = 1|H_1)$, $i = 0, 1, ..., N$ (with P_{F_0} and P_{D_0} representing the false alarm and detection rates of the fusion center).

In the Bayesian formulation, the optimal decision rules minimize the average cost function also called the Bayes' risk function $\Re$, defined for a binary hypothesis testing problem as

$$\Re = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} P(H_j) P(\text{Decide } H_i | H_j \text{is true}), \tag{2.2}$$

where C_{ij} is the cost of making a decision in favor of H_i when H_j is true and $P(H_i)$, $i = 0, 1$ are the a priori probabilities of the hypotheses. Under such a formulation, the

optimal decision rule at the local detectors and at the fusion center is generally of the form of a likelihood ratio test where the incoming observation vector o_j, $j = 1, .., K$ is used to form the likelihood ratio $\Lambda(o)$ which is then compared to a threshold as follows

$$\Lambda(o) = \frac{P(o_1, ..., o_K|H_1)}{P(o_1, ..., o_K|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t \tag{2.3}$$

In the special case when the costs are assumed to be symmetric and set as $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 1$, the Bayes' risk is just the average probability of error. While minimizing the Bayes' risk, the threshold t is a function of the a priori probabilities and chosen costs as shown below

$$t_g = \frac{P(H_0)(C_{10} - C_{00})}{P(H_1)(C_{01} - C_{11})}, \tag{2.4}$$

Minimizing Bayes' risk using a decision threshold as computed in (2.4) requires the knowledge of the prior probabilities which might not be always available. Also, note that for the case of decision fusion, the observations at the local detectors are $o = z$ and the observations at the fusion center are the local decisions $o = [u_i, ..., u_N]$ (see Figure 2.2).

As an alternative, to circumvent the problem of acquiring prior probabilities, the optimization goal might be defined as to maximize the probability of detection such that the probability of false alarm is constrained to be $P_F \leq \alpha$, $\alpha \in (0, 1)$, namely the Neyman-Pearson criterion [19] can be used, where the threshold t is obtained such

the following is satisfied

$$\int_t^{\infty} P(\Lambda(o)|H_0) = \alpha. \tag{2.5}$$

In the case of parallel binary decision fusion, the local decisions u_i are considered as the observations of the fusion center. As the local decisions are discrete Bernoulli random variables, the likelihood ratio $\Lambda(u)$ (now as a function of the local decisions $u = u_1, u_2, ..., u_N$) has a probability mass function and therefore the analogous discrete form of the expression in (2.5) must be used to calculate the threshold t_g.

2.4 Parallel Decision Fusion Survey

Chair and Varshney [20] developed the optimal fusion rule when statistical characteristics of all the distributed sensors are exactly known and fixed, and the local sensor outputs are statistically independent conditioned on the hypotheses. For a binary hypotheses set $H = \{H_0, H_1\}$ with known a priori probabilities $P(H_0)$ and $P(H_1)$ and symmetric costs ($C_{00} = C_{11} = 0$, and $C_{01} - C_{10} = 1$), if the local decisions are represented as u_1, $i = 1, ..., N$, the fusion rule that minimizes the global probability of error is defined as

$$u_0 = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} w_i u_i > w_0 \\ -1, & \text{otherwise,} \end{cases} \tag{2.6}$$

where, $w_0 = log\left(\frac{P(H_0)}{P(H_1)}\right)$ and the optimal weights are given as

$$w_i = \begin{cases} log\left(\frac{1-P_{M_i}}{P_{F_i}}\right), & \text{if } u_i = 1 \\ log\left(\frac{1-P_{F_i}}{P_{M_i}}\right), & \text{if } u_i = -1. \end{cases} \tag{2.7}$$

The result in (2.6) is a closed-form expression of the fusion rule of the local decisions and is essentially a weighted sum of the local sensor decisions being compared to a constant threshold (namely a classical perceptron). The weights w_i, $i = 1, ..., N$ by which the local decisions u_i are scaled in the fusion rule are functions of the error probabilities (false alarm P_{F_i} and miss-detection P_{M_i} rates of the local sensor-detectors) and the constant threshold is a function, in addition, of the a priori probabilities of the hypotheses set.

A variation of distributed detection was studied in [21] where local detectors communicate with each other and fuse the incoming decisions of other detectors locally to produce an updated local decision. No central fusion center is incorporated in the scheme. The conditions for joint optimality of local processors and the fusion center were derived in [22]. This optimization involves solution of coupled nonlinear equations representing the performance of the local detectors and the global fusion center. In essence, this study combines the ideas of local decision optimization in [3] with the fusion rule optimization proposed in [20]. Distributed fusion of correlated local decisions was investigated among others in [23–25]. Krzystofowicz and Long in [26] developed a Bayesian detection model where the local detectors send forward posterior detection probabilities to the fusion center which then uses Bayes' rule to compute the aggregate posterior detection probability. Distributed detection with feedback have been studied among others in [5, 6, 27]. Thomopoulos et al. in [28] provided a general proof that the optimal decision scheme that maximizes the probability of detection at the fusion center for a fixed false alarm rate consists of a Neyman-Pearson test at the fusion center and likelihood ratio tests at the local

detectors. The study in [29] investigated the problem of distributed detection and fusion from an information-theoretic point-of-view. Design of the entire parallel fusion network based on both Bayesian and Neyman-Pearson formulation was considered in [30] and [31]. A more detailed exposition including distributed detection for multiple hypotheses can be found in [32].

2.5 Distributed Detection with Identical Sensors

An interesting case is when the observations at each detector are assumed to be conditionally independent and identically distributed. Contrary to common intuition, even if observations at each detector are identically distributed (given either hypothesis), it is not optimal for each detector to use identical thresholds as shown by counter examples in [3, 33]. The use of different thresholds at each detector makes the PBPO solution of local decision rules intractable especially when the the number of detectors N is large. This is due to the fact that a search over all nonidentical local decision rules is required. If the local decision rules were constrained to be identical, the complexity of the distributed detection problem is drastically reduced. Furthermore [34] showed that when the observations at each detector are independent and are corrupted by identical Gaussian noise, then no optimality is lost if local likelihood ratio tests employ identical thresholds. Chen et al. in [35] showed that local decision rules of the optimum system are almost identical and therefore marginally different from decision rules obtained under identical decision rule constraint. Numerical results in [36] showed that the restriction of identical decision rules leads to negligible loss of performance.

2.6 Person-By-Person Optimization

As mentioned before, optimizing a distributed decision fusion system in its entirety involves deriving the optimal decision rules for both the local detectors and the fusion center. This distributed binary decision fusion system design can be viewed as a *Team Decision Problem* and a general method for seeking such simultaneous optimization is through the Person-By-Person Optimization (PBPO) method. The distributed detection system is viewed as a team of two members. The group of local detectors forms one member and the decision fusion center is the other member. Performance of each member of the team is optimized separately with the assumption that the other members have already been optimized. This approach requires simultaneous solution of nonlinear coupled equations for local detector thresholds and the global fusion rule. Still, the PBPO optimal solution is not guaranteed to achieve the true team optimum [37]. Only under the special condition when the objective function to be minimized is convex, PBPO solution achieves the team optimum. Furthermore, system design equations resulting from PBPO procedure represent necessary and not in general, sufficient conditions to determine the globally optimum solution.

For a N detector binary decision fusion system with non-identical detectors, the PBPO procedure results in N equations, one for each local detector and 2^N equations for the fusion center. A simultaneous solution of these $N+2^N$ nonlinear coupled equations yields the PBPO solution to the binary distributed hypothesis testing problem. The system wide optimization problem using the PBPO procedure was solved under Bayesian framework in [22, 30] and under the Neyman-Pearson criterion in [31]. The

proposed PBPO solutions require simultaneous solution of non-linear coupled equations for decision thresholds, which can become difficult as the number N of local detectors increases.

2.7 Application to Present Work

One of our objectives is to develop an algorithm that provides the optimal local and global decision rules without having to solve a set of coupled nonlinear algebraic equations. Such efforts are not new- the system-wide PBPO optimal solutions were provided under the Bayesian framework in [22] and using the Neyman-Pearson criterion in [31]. However, PBPO solutions are not generally team optimal, and therefore we desire to develop an alternative algorithm that can achieve the team optimal operating point for a distributed binary decision fusion system.

3: OPTIMAL DISTRIBUTED NEYMAN-PEARSON FUSION

We study the parallel binary distributed detection architecture shown in Figure 2.2. N local sensors/detectors observe a phenomenon. Each local detector decides whether to accept one of two binary hypotheses on the observed phenomenon (H_0 or H_1), and it transmits its decision ($u_i = -1$, accept H_0 or $u_i = 1$, accept H_1) over an ideal communication channel to a decision fusion center, where the local decisions are fused to a global decision u_0 as to whether to accept H_0 or H_1. As mentioned in Chapter 2, several of the works particularly [3, 22, 31] that studied the same problem used the PBPO procedure for designing the entire system (local and fusion center decision rules).

We propose an alternate algorithm [38] that avoids the need to solve a set of nonlinear coupled equations, and provides the system wide optimal operating point (local detector and fusion center decision thresholds) of the parallel binary decision fusion system. In the distributed system, both the local detectors and the decision fusion center use the Neyman-Pearson criterion i.e, the algorithm designed here fixes the global false alarm rate and attempts to compute the local detector likelihood ration test thresholds and the global fusion rule that achieve the maximum global detection probability. The principal effort in the design turns out to be to solve for the roots of a certain N^{th} order univariate polynomial. We also compare the performance of our method with the performance of the Person-by-Person Optimization (PBPO)

approach and that of a centralized detection scheme.

3.1 Distributed Neyman-Pearson Decision Fusion

In the distributed decision fusion system shown in Figure 2.2, the i^{th} local sensor/detector decision with $i = 1, ..., N$ and the global fusion center decision ($i = 0$) are of the form:

$$u_i = \begin{cases} 1, & \text{if } H_1 \text{ is decided} \\ -1, & \text{if } H_0 \text{ is decided.} \end{cases} \tag{3.1}$$

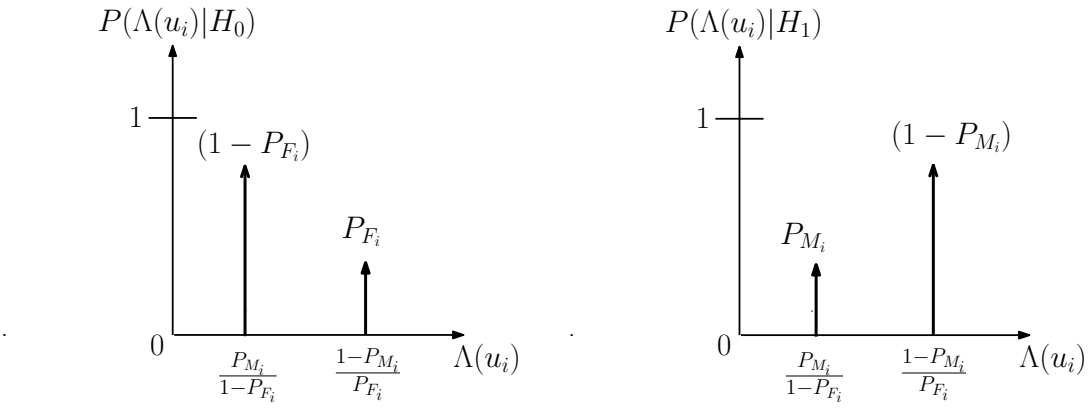
The Neyman-Pearson test fixes the global false alarm rate ($P(u_0 = 1|H_0)$) at a pre-specified level $\alpha < 1$ and then attempts to achieve the maximum global probability of detection ($P(u_0 = 1|H_1)$). The final fusion center decision rule becomes a likelihood ratio test [19], and takes the form

$$\Lambda(u) = \frac{P(u_1, ..., u_N|H_1)}{P(u_1, ..., u_N|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_g. \tag{3.2}$$

The threshold t_g is computed such that the global false alarm is equal to α (Note, we only consider the equality constraint $P(u_0 = 1|H_0) = \alpha$). Assuming that the local decisions are independent (conditioned on the hypothesis), we have

$$\Lambda(u) = \prod_{i=1}^{N} \frac{P(u_i|H_1)}{P(u_i|H_0)} = \prod_{i=1}^{N} \Lambda(u_i) \underset{H_0}{\overset{H_1}{\gtrless}} t_g. \tag{3.3}$$

As the local decisions u_i are binary, the conditional probability distributions $P(\Lambda(u_i)|H_0)$ and $P(\Lambda(u_i)|H_1)$ for the i^{th} sensor are discrete as shown in the Figure 3.1. We use



(a) Under H_0. **(b)** Under H_1.

Figure 3.1: Probability mass function of local detector likelihood for parallel decision fusion.

P_{M_i}, P_{F_i} to denote the *Mis-detection Rate* (MD) ($P(u_i = -1|H_1)$) and *False Alarm Rate* (FA) ($P(u_i = 1|H_0)$) of the i^{th} sensor, respectively.

The problem of system wide optimization of a distributed system inherently is extremely difficult. However, the complexity drastically reduces when the observations at each local detector are assumed to identically distributed. Furthermore, a choice of identical local detector thresholds generate asymptotically optimum solution [35] and numerically the assumption of identical local detectors leads to little or no loss of performance [39], [36].

For our work, we consider a decision fusion system consisting of N identical local detectors with local false alarm and mis-detection rates given respectively as $P_{F_i} = P_F = p$ and $P_{M_i} = P_M = q$, with $p, q \in (0, 1)$, $i = 1, ..., N$. The likelihood ratio $\Lambda(u)$ at the fusion center is computed using discrete Bernoulli random variables (local binary decisions u_i), and the conditional probability mass functions $P(\Lambda(u)|H_0)$ and

$P(\Lambda(u)|H_1)$ can be expressed using the binomial distributions as

$$P(\Lambda(u)|H_1) = \sum_{k=0}^{N} \binom{N}{k} (1-q)^k (q)^{N-k}. \left[\delta \left\{ \Lambda(u) - \left(\frac{q}{1-p} \right)^{N-k} \left(\frac{1-q}{p} \right)^k \right\} \right] \tag{3.4}$$

and

$$P(\Lambda(u)|H_0) = \sum_{k=0}^{N} \binom{N}{k} (p)^k (1-p)^{N-k}. \left[\delta \left\{ \Lambda(u) - \left(\frac{q}{1-p} \right)^{N-k} \left(\frac{1-q}{p} \right)^k \right\} \right], \tag{3.5}$$

where the (Kronecker) delta function $\delta(.)$ is defined as

$$\delta(x) = \begin{cases} 0, & \text{if } x \neq 0 \\ 1, & \text{if } x = 0 \end{cases}$$

In the case of N identical detectors, the distributions in (3.4) and (3.5) will have $N+1$ probability masses. Let us index them by $k = 0, 1, ..., N$. An arbitrary global false alarm probability $P_{F_0} = \alpha$ can be realized as a convex combination

$$\alpha = (1-\gamma) \sum_{k=k'}^{N} \binom{N}{k} (p)^k (1-p)^{N-k} + \gamma \sum_{k=k'-1}^{N} \binom{N}{k} (p)^k (1-p)^{N-k}, \tag{3.6}$$

where k' is the smallest value of $k \in [0, 1, ..., N]$ such that

$$\alpha > \sum_{k=k'}^{N} \binom{N}{k} (p)^k (1-p)^{N-k}$$

and the parameter $\gamma \in [0,1]$ is given by

$$\gamma = \frac{\alpha - \sum_{k=k'}^{N} \binom{N}{k}(p)^k(1-p)^{N-k}}{\sum_{k=k'-1}^{N} \binom{N}{k}(p)^k(1-p)^{N-k} - \sum_{k=k'}^{N} \binom{N}{k}(p)^k(1-p)^{N-k}}. \tag{3.7}$$

The global probability of detection then becomes

$$P_{D_0} = (1-\gamma) \sum_{k=k'}^{N} \binom{N}{k}(1-q)^k(q)^{N-k} + \quad \gamma \sum_{k=k'-1}^{N} \binom{N}{k}(1-q)^k(q)^{N-k}. \tag{3.8}$$

We are looking for the value of the likelihood ratio $\Lambda(u)$ such that the sum of all the probability masses at and to the right of $\Lambda(u)$ is equal to α. For identical sensors, the fusion rule is always *k out of N* [11, 29]. In this case, the desired $\Lambda(u)$ must satisfy the following

$$\Lambda(u) = \left(\frac{q}{1-p}\right)^{N-k} \left(\frac{1-q}{p}\right)^{k} \tag{3.9}$$

for some $k \in [0, 1, ..., N]$. In that scenario, the parameter γ reduces to either 0 or 1. Let k^* denote the value of k that satisfies the global false alarm constraint namely

$$P_{F_0}(k^*) = \alpha = \sum_{k=k^*}^{N} \binom{N}{k}(p)^k(1-p)^{N-k}, \tag{3.10}$$

and the corresponding global probability of detection is

$$P_{D_0}(k^*) = \sum_{k=k^*}^{N} \binom{N}{k}(1-q)^k(q)^{N-k}. \tag{3.11}$$

The system-wide optimal solution is therefore the pair (p, k^*) obtained by solving (3.10) for p for every $k^* \in [1, ..., N]$ and then choosing the pair that maximizes (3.11). Since $\alpha < 1$, $k^* \neq 0$. Noting that for a fixed k^*, the summand in (3.10) is a monotonically increasing function of p, the solution for p in (3.10) in the feasible region of (0,1) is unique. Therefore, if the N roots of the equation

$$\sum_{k=k^*}^{N} \binom{N}{k} (p)^k (1-p)^{N-k} - \alpha = 0 \tag{3.12}$$

are evaluated for every k^*, there would be up to N distinct solutions (one for each k^*). Each one of these solutions would correspond to a value of the global probability of detection (from (3.11)). The optimal local false alarm rate would then be the one that provided the maximum global probability of detection. We illustrate with a simulated example.

Example: Let the local detector observations z_i be of the form

$$z_i = \begin{cases} m + e_i & \text{under } H_1 \\ e_i, & \text{under } H_0, \end{cases} \tag{3.13}$$

where e_i is Normally distributed with zero mean and standard deviation σ ($\mathcal{N}(0, \sigma^2)$) and m is a known constant. As an example, we use here $m = 3$ and $\sigma = 6$. We assume that there are $N = 6$ sensors, and that each sensor collects $K = 5$ observations before making a decision. The global false alarm rate is chosen to be $\alpha = 0.05$. In Figure 3.2 we show the variation of the global probability of detection as the local sensor false

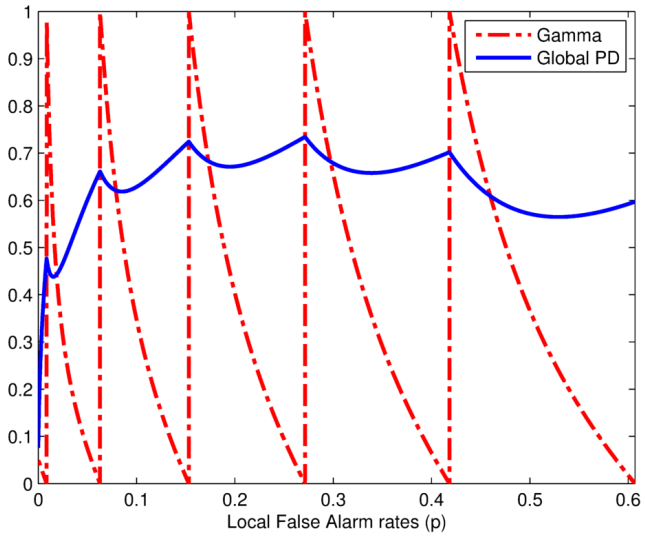


Figure 3.2: Variation of global probability of detection (P_{D_0}) and γ with local sensor false alarm rate (p) for identical sensors.

alarm rate (p) is varied . Note that the x axis runs only till the value of p such that the chosen global false alarm α is greater than or equal to the rightmost probability mass of the conditional distribution $P(\Lambda(u)|H_0)$. It is notable that the curve representing the global probability of detection (blue line) has N cusps at the locations where γ switches between its maximum and minimum values (implying a change in k^*); it is not differentiable there. The maximum P_{D_0} is obtained from the N global detection rates corresponding to the N feasible p values ($p \in (0, 1)$) obtained by solving (3.10) for every $k^* \in [1, ..., N]$.

The computational burden involved with this approach is to compute roots of the N^{th} order univariate polynomial (3.12). We summarize the algorithm [38] in Table 3.1.

In the following section, we briefly discuss the PBPO approach frequently used for a distributed decision fusion system design and then proceed to compare the

Table 3.1: Optimal Distributed Fusion Algorithm

1. For N sensors, consider possible values of k^* in the range $[1, 2, ..., N]$.
2. Solve for the roots of (3.10) for each value of k^*. For each k^* there will be up to N distinct roots. Let the root which is in the feasible region of $[0, 1]$ for a particular k^* be denoted by $p_\alpha(k^*)$.
3. Assuming the local detector observations have a continuous distribution, compute the corresponding local detector threshold (t_{loc}) using $p_\alpha(k^*)$ as follows
$$\int_{t_{loc}}^{\infty} P(\Lambda(u_i)|H_0) = p_\alpha(k^*).$$
4. Compute the corresponding local mis-detection rate $q_\alpha(k^*)$ as follows
$$\int_{-\infty}^{t_{loc}} P(\Lambda(u_i)|H_1) = q_\alpha(k^*).$$
5. For each possible value of k^*, namely $1, 2, ..., N$, compute the global probability of detection $P_{D_0}(k^*)$ using (3.11) with $q = q_\alpha(k^*)$.
6. Find the value of k^* that provided the maximum value of $P_{D_0}(k^*)$.
7. The corresponding $p_\alpha(k^*)$ is the local false alarm for the local sensors that would provide the best global detection rate for the maximum global false alarm of α.

performance of the PBPO approach with the proposed scheme (Table 3.1) through numerical examples.

3.2 Person-By-Person Optimization

A general method for seeking system-wide design of a decision fusion system is through the Person-By-Person-Optimization (PBPO) method. The distributed detection system is viewed as a team of two members. The group of local detectors forms one

member and the fusion center is the other member. Performance of each member of the team is optimized separately with the assumption that the other member has already been optimized. This approach requires simultaneous solution of nonlinear coupled equations for local detector thresholds and the global fusion rule. Still, the PBPO optimal solution is not guaranteed to achieve the true team optimum [37]. For an N detector binary decision fusion system with non-identical detectors, the PBPO solution is obtained by simultaneous solution of $N + 2^N$ nonlinear coupled equations. When the local detectors are identical and the observations at the local detectors are independent conditioned on the hypothesis, the number of equations for PBPO approach under Neyman-Pearson criterion drops down to three (3). Next, we outline the PBPO solution for identical sensors using the Neyman-Pearson criterion.

3.2.1 PBPO-Optimal Local Detector Thresholds

Under the Neyman-Pearson criterion, the global probability of detection is maximized under the constraint that the global false alarm satisfies $P_{F_0} \leq \alpha$. We therefore form the objective function (Lagrangian) to be maximized as

$$F = P_{D_0} + \lambda(P_{F_0} - \alpha), \tag{3.14}$$

where λ is the Lagrange multiplier. Using (3.10) and (3.11) for identical local detectors, we have

$$F = \sum_{k=k^*}^{N} \binom{N}{k} (1-q)^k (q)^{N-k} + \lambda \left[\sum_{k=k^*}^{N} \binom{N}{k} (p)^k (1-p)^{N-k} - \alpha \right]. \tag{3.15}$$

Expanding (3.15) in terms of q and p, the probability of mis-detection and false alarm of a local detector respectively, we have

$$F = (1-q)\sum_{k=k^*}^{N}\binom{N}{k}(1-q)^{k-1}(q)^{N-k} + \lambda\left[p\sum_{k=k^*}^{N}\binom{N}{k}(p)^{k-1}(1-p)^{N-k} - \alpha\right]. \tag{3.16}$$

Let us define the expressions

$$V_p = \sum_{k=k^*}^{N}\binom{N}{k}(p)^{k-1}(1-p)^{N-k},$$

and

$$V_q = \sum_{k=k^*}^{N}\binom{N}{k}(1-q)^{k-1}(q)^{N-k}.$$

The expression in (3.16) becomes

$$F = (1-q)V_q + \lambda(pV_p - \alpha). \tag{3.17}$$

Since V_q is sum of positive real numbers, $V_q \neq 0$. Hence we have

$$F_q = \frac{F}{V_q} = (1-q) + \frac{\lambda V_p}{V_q}(p - \frac{\alpha}{V_p}). \tag{3.18}$$

Maximizing F_q implies that each local detector maximizes its own probability of detection $(1-q)$ subject to the constraint that its local false alarm is bounded as

$p \leq \frac{\alpha}{V_p}$. Each local detector performs a likelihood ratio test as

$$\frac{P(z_i|H_1)}{P(z_i|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_{loc}, \tag{3.19}$$

where z_i are the observations for the i^{th} detector and the local threshold t_{loc} is computed such that the local false alarm is fixed at $p = \frac{\alpha}{V_p}$. In other words, (3.18) becomes the Lagrangian for a local detector. Under the Neyman-Pearson criterion, t_{loc} is given by the Lagrange multiplier [19] - therefore implying $t_{loc} = \frac{\lambda V_p}{V_q}$, where λ is the threshold for the global likelihood ratio test (t_g in (3.2)). From the local detector optimization, we obtain

$$t_{loc} = \frac{\lambda V_p}{V_q}, \tag{3.20}$$

$$p = \frac{\alpha}{V_p}. \tag{3.21}$$

3.2.2 PBPO-Optimal Global Fusion Rule

Since the local detectors are identical, the global fusion rule is a k out of N rule. The optimal k (denoted by k^*) can be obtained by noting that the Lagrange multiplier in (3.14) is effectively the threshold of the global likelihood ratio test or the value of $\Lambda(u)$ at which the global false alarm constraint ($P(u_0 = 1|H_0) = \alpha$) is satisfied, and therefore from (3.9) we have

$$\left(\frac{q}{1-p}\right)^{N-k^*} \left(\frac{1-q}{p}\right)^{k^*} = \lambda. \tag{3.22}$$

Taking natural logarithm of both sides, we have

$$k^* \left[log(\frac{1-q}{p}) - log(\frac{q}{1-p}) \right] = log(\lambda) - Nlog(\frac{q}{1-p}) \tag{3.23}$$

Since the constraint is $P_{F_0} \leq \alpha$, we can express the optimal k^* as

$$k^* = \left\lceil \frac{log(\lambda) - Nlog(\frac{q}{1-p})}{\left[log(\frac{1-q}{p}) - log(\frac{q}{1-p}) \right]} \right\rceil, \tag{3.24}$$

where $\lceil \ \rceil$ is the ceiling function defined over the set of integers (Z) as

$$\lceil x \rceil = min\{s \in Z | s \geq x\}.$$

The complete PBPO solution for identical sensors under Neyman-Pearson criterion therefore requires the simultaneous solution of the coupled nonlinear equations (3.20), (3.21) and (3.24).

In general the PBPO solution does not converge to the team optimum solution. Bauso and Pesenti in [37] showed that the necessary and sufficient condition for a PBPO solution to converge to the team optimum is satisfied when the team cost function has a unique local minimum. This is not the general case for the problem we study, as shown, for example, in Figure 3.2 where the global probability of detection is not unimodal with respect to the local detector false alarm rate, p. Therefore the simultaneous solution of (3.20), (3.21) and (3.24) is not expected to achieve the team optimum solution for all ranges of global false alarm. Furthermore, the PBPO Receiver Operating Characteristics (ROC) is a collection of different ROC curves

(each corresponding to a different value of the optimal k^*); the collective PBPO ROC is formed using the upper envelopes of each of those constituent ROC curves. Due to this feature, even though the ROC curves corresponding to any particular k^* is concave, the overall PBPO ROC curve is not concave as the PBPO optimal k^* changes over various ranges of global false alarm rates. The above mentioned properties are illustrated in the comparative analysis presented in the next section.

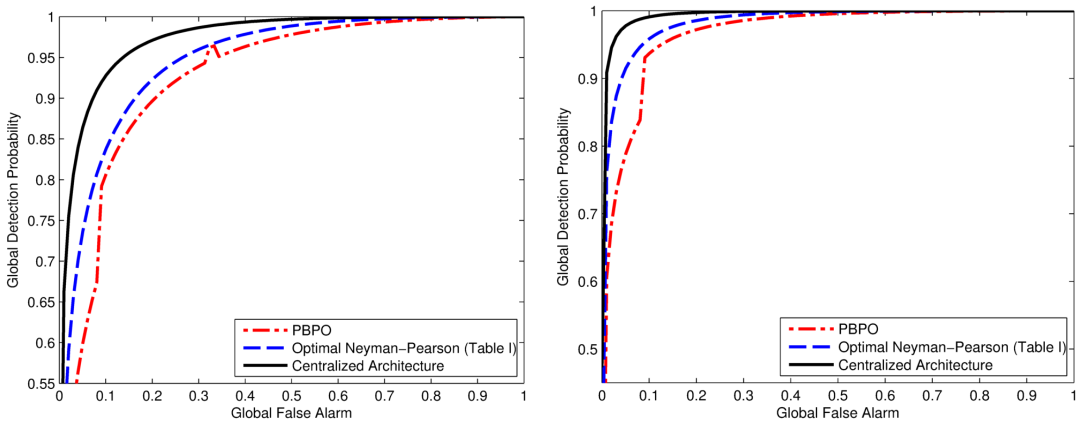
3.3 Examples and Discussion

We provide a performance comparison of our method with the PBPO approach using ROC curves for several scenarios. We also include the performance of a centralized fusion scheme, where the fusion center receives the raw observations and computes the global decision with no involvement of local detectors. Since the centralized architecture performs no local data compression, it provides an upper bound on the performance of a parallel fusion system. In the scenario for the centralized architecture, the fusion center receives $K_c = NK$ observations and uses a Neyman-Pearson test with specified false alarm probability to arrive at a decision. We consider the following three cases:

1. Observations are Gaussian distributed with different means under the two hypotheses, namely:

$$P(z_i|H_1) \sim \mathcal{N}(m, \sigma^2) \tag{3.25}$$

$$P(z_i|H_0) \sim \mathcal{N}(0, \sigma^2). \tag{3.26}$$

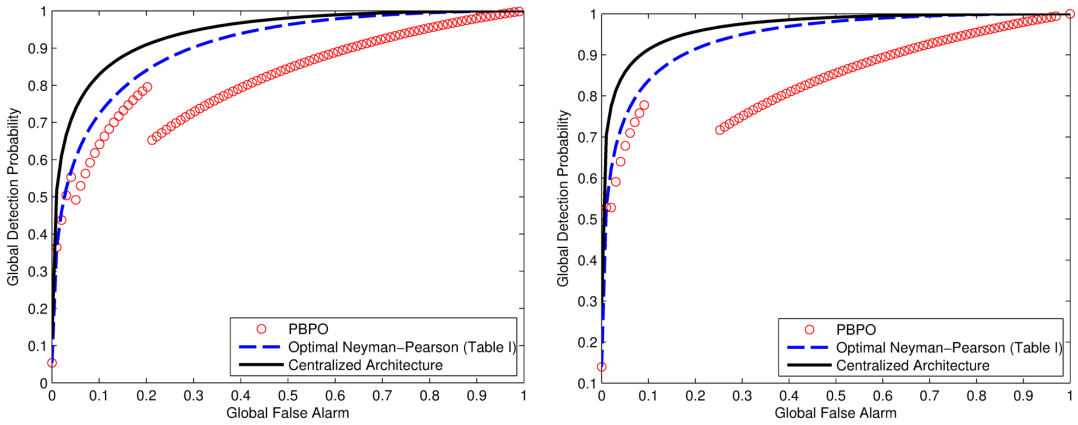


(a) ROC for m = 3 **(b)** ROC for m = 4

(c) ROC for m = 5 **(d)** ROC for m = 6

Figure 3.3: ROC curves under various SNR for distributed Neyman-Pearson detection using optimal distributed fusion algorithm (Table 3.1); Distributed Neyman-Pearson detection using PBPO; and Centralized Neyman-Pearson detection.

where the standard deviation $\sigma = 6$. ROCs are shown in Figure 3.3 for three systems, namely a) Distributed detection using optimal distributed fusion algorithm (Table 3.1); b) Distributed Neyman-Pearson detection using PBPO; and c) Centralized detection. Four different values were used for m (this basically represents varying SNR values) in (3.25), namely 3, 4, 5 and 6. Figure 3.3 shows the extent to which the optimum detection scheme (Table 3.1) improves over



(a) Exponential distribution **(b)** Gamma distribution

Figure 3.4: Performance comparison of the three systems when local detector observations are Exponential and Gamma distributed.

PBPO. The centralized system is of course better than both.

2. Observations are Exponentially distributed,

$$P(z_i|H_1) \sim Exp(1/2) \tag{3.27}$$

$$P(z_i|H_0) \sim Exp(1/3). \tag{3.28}$$

3. Observations are Gamma distributed,

$$P(z_i|H_1) \sim Gamma(1,2) \tag{3.29}$$

$$P(z_i|H_0) \sim Gamma(1,1). \tag{3.30}$$

Figure 3.4 shows the ROC curves for the cases when local observations were Exponential and Gamma distributed, and documents the improvement provided by the

optimal algorithm (Table 3.1). Depending on the distribution of the local detector observations, some values of global false alarm may not have a corresponding PBPO solution. Several such regions are noticed in Figure 3.4.

3.4 Summary and Future Work

We considered system-wide optimization of a distributed decision fusion system where a group of local sensor/detectors perform binary hypothesis testing on observations from a common volume of surveillance, and communicate their decisions to a decision fusion center. The objective is to maximize global probability of detection under a global probability of false alarm constraint. The local detector decision thresholds and the global fusion rule were derived by computing the roots of a N^{th} order polynomial where N is the number of local detectors. The proposed method was compared against the traditional PBPO approach. ROC curves of several scenarios demonstrate the extent to which the optimal solution outperforms PBPO, and the extent to which it is over performed by a centralized detection scheme (where all the raw observations are transmitted to the fusion center).

The major computational burden in the proposed algorithm arises from the requirement of solving for roots of certain N^{th} order univariate polynomials. In practical scenarios, distributed detection systems may consist of large number of sensors N, which may make the proposed algorithm difficult to implement. However, newer root finding algorithms [40, 41] can possibly aid in developing the proposed algorithm further and provide an efficient means of practical implementation.

Furthermore, the proposed algorithm assumes an equality constraint on the global

probability of false alarm ($P_{F_0} = \alpha$). It is not straightforward to generalize the algorithm when the inequality constraint $P_{F_0} \leq \alpha$ is used. Such a constraint may provide optimal ranges for local thresholds and it would be interesting if such bounds could be derived.

Part II

DISTRIBUTED DETECTION WITH HUMANS AS INFORMATION SOURCES

4: EVIDENCE THEORY AND THE HARD/SOFT FUSION PROBLEM

Distributed detection as presented in the previous chapters dealt with *Hard* sensors, in the sense that the sensors/detectors have well defined statistical performance characteristics. However, modern trends in data assimilation and aggregation applications increasingly show that apart from such electronic or mechanical *Hard* sensors, humans too often act as information sources. It is therefore desirable to incorporate their decisions in the fusion scheme. Humans acting as information sources rather than information analysts are termed soft sensors. The task of a distributed detection system involving a heterogeneous mix of hard and soft sensors is to combine local information sources so as to generate a reliable global decision or support value toward one of the hypotheses. The fusion algorithms that can facilitate such *Hard/Soft* combination often require additional tools to those used with hard sensors alone.

4.1 The Hard/Soft Fusion Problem

Figure 4.1 gives a generic representation of the *Hard/Soft* fusion problem. The idea is to bring information from human sources which are sometimes imprecise and uncertain in nature to a form which is compatible for fusion with outputs from hard sensors/detectors.

The first issue in modeling soft sensors is to identify means of knowledge extraction. Broadly, we identify the following two avenues which are closely related and

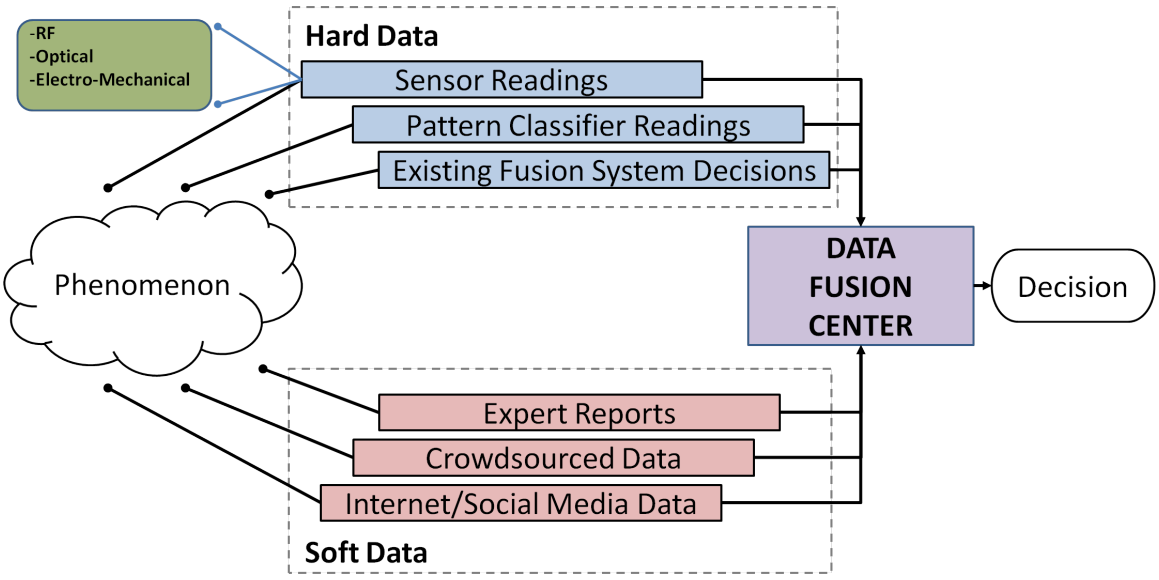


Figure 4.1: The distributed hard/soft fusion scheme.

have sufficient overlap.

- **Indirect**: The indirect method essentially deals with the extraction of information from texts or database systems. It requires tools like natural language processing, text parsing or data mining. Some possible and popular sources may include
  - Global Positioning System (GPS) annotated audio and video
  - Social media data: Facebook, Twitter
- **Direct**: This method gathers information by direct evaluation of human opinions toward a fix set of outcomes or a crisp hypothesis set. Data are gathered using surveys, polls or crowd sourcing.

Here we deal only with direct method of information extraction from soft sources.

4.2 Soft Sensors and Imperfect Data

Once the knowledge elicitation method has been fixed, the next most important challenge in handling soft data is the difficulty in modeling and representing the imperfectness. Imperfect data can be characterized as being imprecise, uncertain, or both. Additionally there are other imperfections like vagueness or incompleteness. The qualitative problem is how to choose a suitable framework to represent and characterize such imperfect information as there is no general formalism to describe all kinds of imperfect data. Once a generic framework for data representation is chosen, the quantitative aspect of the problem asks how to quantify the soft source's confidence accorded to a realization (for example: how the belief of an event is set to be 0.8 and not 0.6).

4.2.1 Types of Imperfect Data

There are various ways of defining classes of imperfect data. The distinction between such classes is not strict and there are overlaps. Some of the most used classifications of imperfect data are the following:

- **Subjective and Objective data**: The subjective information is an information which is dependent on the source of information. This means that the source of information has given an interpretation of the data. Usually, humans are the main contributors of such information. The information is subjective due to the limitations of language or to the limitations of understanding.

- **Uncertain and Imprecise data**: Uncertainty represents our state of knowl-

edge about an item of information. Imprecision is a characteristic of an information that is expressed often by a set of values. For example, "The detected object is a F-18 plane" is a precise and certain information. However, "The detected object might be a F-18 plane with a degree of confidence of 60%" is an example of an uncertain information as the information is precise but not sure. The information "The detected object is one of the planes {F-18, Boeing 747, F-16, Mig-29}" is an imprecise information. Our state of knowledge is that the detected object is one of the element of the set but we are unsure about the exact one. Usually, the information provided by a general soft source is both imprecise and uncertain and the framework chosen to represent such data therefore should be able to handle both of them.

- **Vague and Incomplete data**: This kind of data is due specially to the limitation of the vocabulary and is in most of the cases subjective data. We refer to information which describes a class of objects, but the limits of this class are not well known. For example, "The observer spotted a reconnaissance plane flying very high" is vague, because of the lack of precision in "high". Incomplete data are a combination of imprecise and uncertain information. Incomplete data are represented by the upper limit of our degree of confidence (also called the possibility) of an event.

4.2.2 Frameworks for dealing with Imperfect Data

Research over the past three decades have resulted in a number of theories which aid in representation of imperfect data. In a typical application, the variety of sources

provides different kinds of imperfect data and no theory is universal in applicability (capable of handling uncertain, vague or imprecise information). Some of the most popular theories are listed below. In the subsequent discussions, we will denote the set of all hypotheses by Θ purely due to notational compatibility with the common form of the mathematical expressions in evidence theory. This set basically carries the same meaning; that of a set of disjoint hypotheses as was denoted by H in Part I.

- **Probability Theory**: The probability theory has been the most popular framework to be used to deal with almost all kinds of imperfect information, partly because it was the only existing theory. The measure of probability expresses the degree of confidence that someone assigns to the occurrence of a realization of an event. If $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ denotes the set of possible mutually exclusive realizations/hypotheses, then probability theory assigns precise probability numbers to each member of Θ. In reality such assignment is hardly possible since no one knows the chances of occurrence of an event with 100% accuracy. Probability theory enables fusion of information coming from various sources by using the expression of total probability.

$$P(A) = \sum_{i=1}^{N} P(A|source_i)P(source_i)$$

where the A is the event under consideration. When the prior probabilities are unknown or there is no reliable information about them, probability theory assigns equal probability to all elements of Θ. This is the way to model ignorance in probability theory. This theory can also deal with imprecision, but the prob-

ability of an imprecise event is strongly dependent of the probabilities of precise events. In a lot of cases, the prior information is not available and the user does not have all the data to solve the problem. Moreover, imperfect information, especially the imprecise one, is hardly modeled with the probability theory. For this kind of imperfect information, other theories were proposed.

- **Dempster-Shafer Theory of Evidence**: The Dempster-Shafer's theory of evidence [1] (henceforth we will refer to this theory as Evidence theory) was introduced by Dempster in [42] and later developed by Shafer in [16] as an alternative to the probability theory to fulfill the need of dealing with both imprecision and uncertainty in observed data. Confidence values are associated to elements of the power set 2^{Θ} instead of Θ as in probability theory. This helps in modeling inherent ignorance and uncertainty of the sources. The theory can handle incomplete information and also provide an upper and lower bound on the likelihood of an event. Fusion of data from various sources can be done using Dempster's rule of combination.

- **Rough Sets**: The rough sets theory [43] was proposed to deal with the aspect of imprecision. The basic concept of the rough sets theory is to replace an uncertain imprecise information by two imprecise but certain information: the lower and upper approximations. The combination of imprecise information is realized by applying set theory to the approximations. The salient feature of the rough sets theory is that there is no need to quantify the information's

[1] In literature different naming conventions like Dempster Shafer Theory, Evidence Theory, Belief Theory have been used to refer th this theory

uncertainty.

- **Fuzzy Sets**: Zadeh in [44] developed the idea of fuzzy sets which is based on the concept of uncertain membership to a set. If X is an uncertain subset of Θ called also a fuzzy set then, its membership function is $\mu_X(\theta_i) \in [0, 1]$, where $\mu_X(\theta_i)$ describes the degree of membership of θ_i to X. The theory of fuzzy sets deals especially with vague, ill defined, or ambiguous data.

- **Possibility theory**: The possibility theory [45, 46] is a tool able to deal with both imprecision and uncertainty. For every $\theta_i \in \Theta$ is defined a possibility measure, which represents a limit of the degree of confidence given to it. The possibility measure, $\pi(\theta_i)$, takes values in the [0,1] interval and $\pi(\theta_i) = 1$ means that the event is possible, but does not mean that it is also certain.

The reason why there are multiple frameworks is that there are different types of uncertainty and a better posed underlying theory was not discovered yet. In this thesis, the term "uncertainty" generally refers to *epistemic* uncertainty because it corresponds to beliefs held by an observer or human source about the world. For dealing with *aleatory* uncertainty related to randomness and chance, probability theory is usually the preferred framework. Lack of information may lead to uncertainty which is the result of ignorance rather than randomness. The Bayesian view is that ignorance can be adequately represented using probability theory by applying equal probability to all possible hypotheses. In contrast, belief function theory distinguishes between these types of uncertainty and thus makes ignorance explicit.

4.3 Evidence Theory and Belief Model

Belief function theory was originally developed by Shafer in [16]. It was designed keeping in mind the solution to the inherent difficulties of probability theory in working with imperfect data. Central to the theory is the notion of evidence and how different pieces of evidence should be combined in order to make inferences. Belief function theory can be interpreted as a generalization of Bayesian probability theory.
Example: Consider the situation when an observer is offered a bet on the outcome of a coin toss. Without any prior knowledge, there is no reason to trust that the coin is fair. What should the observer's belief about the possible outcomes be in this state of total ignorance? In the Bayesian framework, both the outcomes are modeled as equiprobable with $P(Heads) = P(Tails) = 0.5$. The Bayesian belief in the case of total ignorance is therefore equivalent to the situation where the coin has been tested extensively and is determined to be fair.

In contrast, a belief function can explicitly represent the state of ignorance by assigning a belief mass to the total frame $m(\Theta) = m(Heads, Tails) = 1$. Such a belief mass assignment abstains from assuming any true probability distributions of the states. This is the critical difference between Bayesian probability and the belief model and bodes very well in representing vague and ignorant opinions of human sources/observers.

Evolution of Belief Theory A detailed account of the historical development of belief function theory is given in [47]. The theory developed by Shafer [16] builds on previous works by Dempster on upper and lower probabilities [42]. There have been

long debates between proponents of Bayesian framework and belief theory about the question whether Bayesian probability theory is sufficient for modeling uncertainty or whether belief function theory is more appropriate [48]. However, many of the arguments are more of a philosophical nature and deal with the rational interpretation of belief and probabilities. There exist multiple schools of thought about the interpretation of beliefs and probabilities, comparisons of which can be found in [49, 50].

Dempster's original work [42] on one-to-many mappings applied to a probability space leads to lower and upper bounds of probabilities and thus constitutes a probabilistic interpretation of belief. In contrast, Shafer's approach in [16] can be interpreted as being non-probabilistic because the framework defined by Shafer has its own axioms and is not derived from probability theory. However, the existence of some partially-known probability measure corresponding to a belief function is usually still assumed. The transferable belief model (TBM) [49, 51, 52] developed by Smets rejects the notion of an underlying probability measure altogether. The TBM consists of two levels: a *credal* level where beliefs are represented and combined, and a *pignistic* level where decisions are made based on probabilities derived from the credal level. A host of seminal works in the field of belief theory and its applications to expert systems can be found in [53].

4.3.1 Elements of Belief Model

The first step in applying the belief model is to define a finite set of mutually exclusive and exhaustive possible states or hypotheses called a *Frame of Discernment* that the event under observation can take. The frame of discernment helps form a set of all

feasible outcomes of the event and each such outcome is referred to as a *proposition*. If Θ is used to define such a set of disjoint hypotheses and forms the frame, then the power set of Θ represented as 2^{Θ} is sometimes an useful choice as the set of propositions. In that case, all elementary states of the set of propositions will be called *atomic* sets, as they do not have any subsets. For example: if $\Theta = \{\theta_1, \theta_2, \theta_3\}$ is a set of mutually exclusive elements forming the frame, then the power set would look like $2^{\Theta} = \{\theta_1, \theta_2, \theta_3, \{\theta_1, \theta_2\}, \{\theta_2, \theta_3\}, \{\theta_1, \theta_3\}, \Theta\}$.

The next step is for an observer or human source to assign belief masses to the various propositions which are members of 2^{Θ}. In this step lies the most significant difference between Dempster-Shafer theory and the Bayesian approach since unlike the latter, the belief masses in Dempster-Shafer theory are assigned to elements of 2^{Θ} and not to Θ. Hence, Dempster-Shafer theory allows us to deal with evidence toward not only the singleton elements but composite sets like $\{\theta_i \cup \theta_j\}$, $\forall \theta_i \in \Theta$ as well. The belief mass assignment is basically associating numbers $m(X) : 2^{\Theta} \mapsto [0, 1]$ to each subset $X \in 2^{\Theta}$ such that they obey the following constraints.

$$m(X) \geq 0 \tag{4.1}$$

$$m(\phi) = 0 \tag{4.2}$$

$$\sum_{X \in 2^{\Theta}} m(X) = 1. \tag{4.3}$$

where ϕ is the null set. Any subset X of Θ such that $m(x) > 0$ is called a *focal element*. Ignorance in evidence theory is modeled by granting a nonzero mass to the set Θ. When $m(\Theta)$ is zero, the data can be said to be not ignorant but can still

be imprecise. Some critical points of differences between evidence theory and Bayes' probability theory manifest when the above expressions are compared with the basic axioms of probability. These distinctions are referred below:

In evidence theory it is **not** required that

- $m(\Theta) = 1$
- $m(X) \leq m(Y), if X \subset Y$
- there be a relationship between $m(X)$ and $m(\overline{X})$

Furthermore, any belief mass assigned to a composite set is allowed to seep down to its constituent subsets in an unknown manner. For instance a mass assigned by a source to a composite set $\{\theta_1, \theta_2\}$ would be inferred as the confidence of the source toward the occurrence of either θ_1 or θ_2 but the source is uncertain as to which one. This is the primary reason why evidence theory can deal with uncertainty of sources of information better than probability theory. The belief mass assigned to a proposition X can be thought of as the proportion of all relevant and available evidence that supports the truth of X. It does not however represent belief in the subsets of X. The *belief function* on the other hand is generally the total belief that is committed to a particular proposition (including all its subsets). The belief function $b(.) : 2^{\Theta} \mapsto [0, 1]$ of a particular proposition X is defined as

$$b(X) = \sum_{A \subseteq X} m(A), \quad A, X \in 2^{\Theta} \ . \tag{4.4}$$

In contrast to Bayesian probability theory, in belief function theory, additivity of

beliefs is not required. This means

$$b(X) + b(\overline{X}) \leq 1, \forall X \subseteq \Theta. \tag{4.5}$$

If the belief masses are normalized a direct consequence of (4.4) is

$$b(\Theta) = 1. \tag{4.6}$$

Another term called plausibility, $pl() : 2^{\Theta} \mapsto [0, 1]$ is defined as

$$pl(X) = \sum_{A \cap A \neq \phi} m(A), \quad X, A \in 2^{\Theta} \ . \tag{4.7}$$

The quantity $b(X)$ can be interpreted as the cumulative evidence toward the truth of the proposition X. The plausibility $pl(X)$ can be viewed as the total evidence that does not contradict X. Note that functions $m(X), b(X), pl(X)$ are in one-to-one correspondence and can be seen as three facets of the same piece of information. From the definitions, some simple properties about belief and plausibility are as follows:

$$b(X) \leq pl(X) \tag{4.8}$$

$$b(\phi) = pl(\phi) = 0 \tag{4.9}$$

$$b(\Theta) = pl(\Theta) = 1 \tag{4.10}$$

$$pl(X) = 1 - b(\overline{X}) \tag{4.11}$$

$$b(X) + b(\overline{X}) \leq 1 \tag{4.12}$$

4.3.2 Special Classes of Belief Mass Assignments

Based on the definition of belief masses (4.3), there are certain classes of belief mass assignments that deserve special mention.

1. **Categorical Belief Mass**: A belief mass assignment is called *categorical* if it is normalized and has only one focal element.

$$m(X) = 1, \quad X \subseteq \Theta \tag{4.13}$$

   The belief mass assignment is called *dogmatic* if the focal element X is a strict subset of Θ.

2. **Simple Support Belief Mass**: A normalized belief mass assignment is called *simple support* if it has at most two focal elements, one of them being the frame of discernment.

$$m(X) = s, X \subseteq \Theta, 0 < s < 1 \tag{4.14}$$

$$m(\Theta) = 1 - s. \tag{4.15}$$

3. **Vacuous Belief Mass**: A belief mass is called *vacuous* if it is categorical and the focal element is the frame of discernment. A vacuous belief mass assignment represents complete ignorance about the set of alternatives.

$$m(\Theta) = 1. \tag{4.16}$$

4. **Consonant Belief Mass**: A belief mass assignment is called *consonant* if the focal elements form a strict hierarchy (are nested). In other words, for focal elements $X_1, X_2, ..., X_j$ of a consonant belief mass assignment the following must be satisfied for some combination,

$$X_1 \subseteq X_2 \subseteq ... \subseteq X_j, \quad X_i \in \Theta. \tag{4.17}$$

As a result of the above definition, consonant belief mass assignments have $|\Theta| + 1$ focal elements (where $|.|$ is the cardinality of a set).

5. **Bayesian Belief Mass**: A *Bayesian* belief mass assignment is normalized and has only atomic sets (subsets of the frame with cardinality 1) as focal elements.

$$\begin{aligned} &m(X) > 0, \quad \forall X \in 2^{\Theta} : |X| = 1 \\ &m(X) = 0, \qquad \text{otherwise.} \end{aligned} \tag{4.18}$$

Evidently, such a belief mass assignment satisfies the Kolmogorov axioms [54] and therefore such belief masses become same as Bayesian probabilities with the following result

$$b(X) = pl(X) = \text{probability of}(X). \tag{4.19}$$

As long as the belief mass assignment does not satisfy (4.18), the belief and plausibility can be interpreted as lower and upper bound of an unknown probability function

(defining the likelihood of the occurrence of a proposition). This interpretation although is only applicable if the belief model is built using Shafer's approach with the assumption of an underlying probability distribution as opposed to Smets' TBM framework. The underlying probability of a proposition X, $P(X)$ is related to belief and plausibility functions by

$$b(X) \leq P(X) \leq pl(X).$$

4.3.3 Source Reliability and Discounting

Belief theory provides an operator called the *discounting* operator that can be used to scale belief mass assignments based on the reliability of an information source. Reliability, here is defined as source's tendency to assign the correct outcome full belief. This can be estimated from simulations over training sets collected from the source. The idea of reliability of a source is analogous to the concepts of probabilities of detection and false alarm used to define the performance of hard sensors/detectors. Let $r_i \in [0, 1]$ represent the reliability of the i^{th} source. The discounting operation produces the discounted belief mass assignment $m_i(X; r_i)$ from the original source belief mass assignment $m_i(X)$ for each $X \subseteq \Theta$ where

$$m_i(X; r_i) = \begin{cases} r_i m_i(X), & X \subseteq \Theta, \\ r_i m_i(X) + (1 - r_i), & \text{otherwise.} \end{cases} \tag{4.20}$$

4.3.4 Belief Combination

In any inferencing or decision making problem, belief functions representing different pieces of evidence need to be combined in a meaningful way. This is why combination rules are a major building block of belief function theory. Typically, each piece of evidence is represented by a separate belief function on the same frame. Combination rules are then used to successively fuse all these belief functions in order to obtain a belief function representing all available evidence. The combination rule organic to the belief theory proposed by Shafer is referred to as the Dempster's rule ([16], Chapter. 3) and is arguably the most important among all belief combination rules. Assuming two independent belief mass assignments m_1 and m_2 from sources S_1 and S_2 for the same frame, the combined mass $(m_1 \oplus m_2)$ obtained by the Dempster's rule of combination is computed as

$$\begin{aligned} (m_1 \oplus m_2)(X) &= \frac{\sum_{A \cap B = X} m_1(A) m_2(B)}{1 - K}, x \neq \phi \\ (m_1 \oplus m_2)(\phi) &= 0 \end{aligned} \tag{4.21}$$

where, the K is the normalizing factor given by

$$K = \sum_{A \cap B = \phi} m_1(A) m_2(B) \tag{4.22}$$

K represents belief mass associated with conflict or a measure of the disagreement between the sources of information S_1 and S_2. It is determined by summing the products of the basic mass assignments of all sets where the intersection is null. The denomina-

tor in (4.21) assures that the resulting combined mass satisfies $\sum_X (m_1 \oplus m_2)(X) = 1, X \in 2^\Theta$. The more normalization required, the more conflicting the evidence and when $K = 0$, no normalization is required. Dempster's rule cannot be used when $K = 1$. In other words when the sources are completely in disagreement, (4.21) fails. Normalizing with the factor $1 - K$ has the effect of completely ignoring conflict and attributing any belief mass associated with conflict to the null set. Algebraically the Dempster's rule is both associative and commutative. This can be represented as

$$\text{Commutativity:} \qquad m_1 \oplus m_2 = m_2 \oplus m_1.$$

$$\text{Associativity:} \qquad m_1 \oplus (m_2 \oplus m_3) = (m_1 \oplus m_2) \oplus m_3.$$

The normalizing step which essentially disregards conflict between constituent belief mass assignments might result in logically unintuitive outcomes in certain situations. This was brought forward in the famous criticism of Dempster's rule by Zadeh through the example shown in [55, 56]. A considerable research effort since Zadeh's example has resulted in a jungle of combination rules each having its advantages and weaknesses. A moderately exhaustive survey of belief combination rules can be found in [57]. Most of the new combination rules are variations of Dempster's rule and only differ in how they manage the conflicting evidence.

Conjunctive Combination Rule: The conjunctive combination rule (denoted by $\odot$) is a modified version of Dempster's rule and was first proposed in the *Transferable Belief Model* (TBM) framework [51, 52]. Since the TBM framework explicitly allows unnormalized belief functions, the normalization step performed by Dempster's rule is

omitted. Otherwise it is identical to Dempster's rule. For two belief masses (possibly unnormalized) the conjunctive combination rule is defined as

$$(m_1 \odot m_2)(X) = \sum_{A \cap B = X} m_1(A) m_2(B), X \subseteq \Theta. \tag{4.23}$$

Contrary to Dempster's rule the conjunctive rule in (4.23) is always defined and also allows non zero belief masses to be assigned to the null set ϕ. For example, in the case of totally conflicting belief masses the conjunctive combination rule would produce a combined belief mass of $(m_1 \odot m_2)(\phi) = 1$.

Some prominent combination rules based on the conjunctive operation as defined in (4.23) are the Yager's rule [58], Weighted Average Operator (WAO) [59], Consensus Operator [60, 61] and the suite of rules called Proportional Conflict Redistribution (PCR) rules developed in [62, 63].

Disjunctive Combination Rule: The disjunctive combination rule [64] (denoted by $\otimes$) is applicable when only one of the several pieces of evidence holds. In other words, whereas Dempster's rule and the conjunctive combination rule act like a logical AND operation where all the distinct pieces evidence on which the constituent belief masses are based off are considered as possible, disjunctive rule acts like a logical OR operation. For two belief masses (possibly unnormalized) the disjunctive combination rule is defined as

$$(m_1 \otimes m_2)(X) = \sum_{A \cup B = X} m_1(A) m_2(B), X \subseteq \Theta. \tag{4.24}$$

As the set $A \cup B$ is never empty unless both focal elements are individually empty, there is no conflict resulting from the disjunctive rule and hence no need for normalization. Some prominent disjunctive combination rules are Dubois and Prade's rule [64, 65].

4.3.5 Belief Space and Decision Space

Belief theory is a method of specifying imprecise evidence that results in *classes* of subjective probabilities (i.e., belief and plausibility intervals). To facilitate decision making, *probability transformations* are used to map such subjective probabilities in the belief space to a probabilistic measure in the decision space which satisfies the Kolmogorov axioms and that can be used to infer a decision about a particular hypothesis. As described by Smets [49, 66], probability transforms map from the credal to the pignistic level. The *pignistic* transformation proposed by Smets [66, 67] has been one of the first and most popular transformations. For a belief mass m associated to a frame Θ, the pignistic transformation $BetP(.) : \Theta \mapsto [0, 1]$, $\forall X \in \Theta$ is defined as

$$BetP(X) = \sum_{A \subseteq \Theta A \neq \phi} \frac{|X \cap A|}{|A|} m(A). \tag{4.25}$$

Several other such probability transforms have been proposed such as [60, 68–70].

4.3.6 Computational Complexity

One of the major criticisms directed against belief function theory is its computational complexity. Since belief masses can be assigned to arbitrary subsets of the space under consideration (belief masses are assigned to elements of 2^{Θ} instead of

Θ), the complexity of representing and combining belief functions is exponential in the worst case [71]. The larger the cardinality of the frame, more is the number of computations for belief combination. Couple of strategies for reducing such computational complexity associated with belief functions would be to use frames with small focal sets or to exploit independence of evidence and thereby frames defined on such evidence.

4.4 Application of Evidence theory to Present Work

As discussed in the previous chapters, many optimal and suboptimal fusion methods exist for fusion of data from hard sources. However, very few frameworks have so far been developed for Hard/Soft fusion. Furthermore, there are no working examples that test such frameworks on practical scenarios. One of the goals of this thesis is to develop frameworks which can facilitate Hard/Soft fusion and test such algorithms against realistic situations.

The Dempster-Shafer or evidence theory framework has proven itself to be an extremely versatile tool to deal with imperfect data. This study specifically deals with how evidence theory can be used toward problems concerning combination of imperfect data from soft sources and how such schemes can be helpful for Hard/Soft data fusion. We propose evidence theory based architectures which are versatile to receive and fuse both and hard and soft data for inferencing and decision making. Specifically we use the combination rules and probability transformations proposed in [60] to build a belief theoretic framework for modeling soft information sources. Data from such sources are then combined with hard sensors for final decision making. We

present the relevant details in the following chapter.

We also develop new, hierarchical evidence tree based frameworks that can handle complicated and equivocal human statements about a certain hypothesis set.

5: HARD/SOFT FUSION USING BELIEF CALCULUS

The Hard/Soft fusion problem was described in Chapter 4 (see Figure 4.1). The goal is to mathematically model subjective and inherently uncertain data from human sources so as to transform such data to forms compatible for fusion with hard sensor data. The Hard/Soft fusion problem with the so called *human in the loop* situation is new and therefore challenging as compared to the classical problem of distributed detection with physics based hard sensors with known or estimable statistical error characteristics. Current state of the art in the field of data fusion lacks any substantial work facilitating such Hard/Soft data integration [72]. Couple of frameworks and approaches [2, 73, 74], were proposed but had not been actually implemented and verified with simulated or real data. Some studies have tested algorithms which mainly deal with the indirect method of knowledge elicitation (see Section 4.1) and caters to very specific data sets [75, 76].

Here we present the study in [77], where we developed a framework based on techniques in evidence theory, that can be used as an easy to implement hard and soft fusion algorithm. Through a variety of realistic examples, the algorithm is shown to be capable of integrating data provided by hard sensors like GPS sensors and satellite imaging devices with data provided by soft sensors like reports from humans or context analysis by domain experts. Toward this end, we form a probabilistic representation of soft sensor data using evidence theory's belief mass assignment and

a Consensus operator [60], for combining human opinions with uncertainties. We then use the probability fusion rule proposed in [26] to form a hard and soft data fusion system. This approach brings all sensor outputs to the same probabilistic framework prior to fusion and therefore eliminates the inherent differences in hard and soft sensor data types. This work was one of the first that showcased and tested an implementable Hard/Soft fusion architecture. Before illustrating the algorithm, we provide some necessary background material.

5.1 Consensus Operator and Probability Expectation

Dempster's rule of combination (4.21) has difficulty of dealing with contradicting belief mass assignments from multiple sources. The conjunctive combination rule (4.23) can handle all degrees of contradiction but assumes an *open world* and therefore allows a nonzero belief mass assignment to the null set ϕ. The disjunctive combination rules (4.24) performs a logical OR like operation and hence may lead to bias in the combined result. There are many other rules in the above mentioned categories all of which handle uncertainty in different manners. The Dempster's rule (4.21) normalizes and in effect filters out the conflict, redistributing it to all focal elements of the power set. Other rules like Yager's rule [58] assigns the conflict to the frame and thus increases ignorance. The Consensus operator which was proposed as the belief combination rule under the theory of Subjective Logic developed in [61] performs an averaging operation and can handle conflicts in a manner which is intuitively more appealing. Under circumstances where there is a set of conflicting belief mass assignments, the Consensus operator appears to be a better choice for combining such

data (see example in [60]). In cases where there is no conflict, Dempster's rule and the Consensus operator perform almost similarly.

5.1.1 Opinion Tuple

A new metric called *'opinion'* forms the basic building block of the consensus operator. The opinion ω is a tuple consisting of 4 pieces of information, namely the belief (b), disbelief (d), uncertainty (u) and relative atomicity (a). The opinion about a proposition X basically has the form

$$\omega_X = (b(X), d(X), u(X), a(X)). \tag{5.1}$$

We have already defined the belief function b in (4.4). The disbelief in a proposition X is the belief in $\overline{X}$ and is defined as

$$d(X) = \sum_{A \cap X = \phi} m(A), \quad X, A \in 2^{\Theta}. \tag{5.2}$$

The uncertainty for a given set X is expressed as the sum of the belief masses on supersets or on partly overlapping sets of X.

$$u(X) = \sum_{A \cap X \neq \phi, A \not\subseteq X} m(A), \quad X, A \in 2^{\Theta}, X \neq \phi. \tag{5.3}$$

As shown in [61], the following holds:

$$b(X) + d(X) + u(X) = 1, \quad X \in 2^{\Theta}.$$

Based on the definitions of belief and plausibility shown in (4.4),(4.7) the following can be inferred

$$pl(X) = 1 - d(X), \tag{5.4}$$

$$pl(X) - b(x) = u(x). \tag{5.5}$$

The fourth co-ordinate of the opinion tuple is called the *relative atomicity* and is denoted by $a()$. For any particular set X, the atomicity of X is the number of atomic sets it contains or basically its cardinality, denoted by $|X|$. In [61] the relative atomicity of a proposition X relative to another proposition Y is defined as

$$a(X|Y) = \frac{|X \cap Y|}{|Y|}. \tag{5.6}$$

5.1.2 Probability Expectation

In the framework of Subjective Logic, the credal or belief level is mapped to the pignistic or decision level through the transformation *Probability Expectation*. For a frame of discernment Θ, the probability expectation represented as $Pe(X) : 2^{\Theta} \mapsto [0, 1]$, is defined by

$$Pe(X) = \sum_{Y} m(X)a(X|Y), X, Y \in 2^{\Theta}. \tag{5.7}$$

Note that the above definition (5.7) is similar to the construction of the expression of total probability in (4.2.2). Therefore, an analogy can be drawn by comparing relative

atomicity with conditional probability. In (5.7), the term $m(Y)$ is the belief mass assigned to set y and the term $a(X|Y)$ denotes the amount of confidence toward the truth of set X that can be generated if the set Y is true. The probability expectation $Pe(X)$ for any particular proposition X is a mapping of the 4 tuple opinion ω on to the probability space [0 1] and is the expected probability of X being true (note that the expectation is not in the statistical sense but based on the evidence and belief mass assignments) such that

$$\sum_i Pe(X_i) \quad = \quad 1, X_i \in 2^{\Theta}. \tag{5.8}$$

In general, the relation in (5.8) holds only when the sets of propositions are the singletons. In case focal elements have composite sets as well, a normalization might be required [78].

The definition in (5.7) is equivalent with the pignistic probability described in [51], and is based on the principle of insufficient reason; A belief mass assigned to the union of n atomic states is split equally among these n states. The probability expectation of a given state is thus determined by the belief mass assignment and the atomicities. It should be noted that the probability expectation function removes information and that there can be infinitely many different belief mass assignments that correspond to the same probability expectation value.

5.1.3 Focused Frame of Discernments

In order to simplify the representation of beliefs for proposition sets, we will use *focused frame of discernments* which will always be binary, i.e., it will only contain (focus on) one particular set and its complement. More precisely it can be defined as follows. Let Θ be a frame of discernment and let $X \in 2^{\Theta}$. The frame of discernment denoted by $\widetilde{\Theta^X}$ containing only X and $\overline{X}$ where $\overline{X}$ is the complement of X in Θ is then called a *focused frame of discernment with focus on* X. Josang in [61] showed that the focused frame of discernment and the corresponding belief mass assignment will for the proposition in focus produce the same belief, disbelief and uncertainty functions as the original frame of discernment and its associated belief mass assignment. In other words, if Θ is the original frame with $b(X)$, $d(X)$ and $u(X)$ as the belief, disbelief and uncertainty functions for proposition X then in the focused frame $\widetilde{\Theta^X}$ with focus on X (i.e. containing X and $\overline{X}$), the belief mass assignment is defined as follows

$$m_{\widetilde{\Theta^X}}(X) = b(X), \tag{5.9}$$

$$m_{\widetilde{\Theta^X}}(\overline{X}) = d(X), \tag{5.10}$$

$$m_{\widetilde{\Theta^X}}(\widetilde{\Theta^X}) = u(X). \tag{5.11}$$

The focused relative atomicity of the proposition X is however defined as

$$a_{\widetilde{\Theta^X}}(X) = \frac{Pe(X) - b(X)}{u(X)} \tag{5.12}$$

where *Pe(X)* is calculated using (5.7). The focused relative atomicity is defined as in (5.12) so that the probability expectation value of proposition X is equal in both the original and the focused frames. Hence, instead of working with a complex set of propositions, the focused frame which by definition is binary, allows a much simpler and tractable way to implement the consensus operator.

5.1.4 Consensus Operator

Based on the concepts explained above, the consensus operator can be defined as a mathematical framework which can be used to combine multiple opinion vectors each owned by a separate soft sensor. This results in a fused opinion tuple which is the aggregate opinion about the particular proposition. In [61] the consensus operator for combining two opinions $\omega_X^{S_1} = (b_X^{S_1}, d_X^{S_1}, u_X^{S_1}, a_X^{S_1})$ and $\omega_X^{S_2} = (b_X^{S_2}, d_X^{S_2}, u_X^{S_2}, a_X^{S_2})$ respectively owned by soft sensors S_1 and S_2 about a proposition X, is defined as follows.

$$\begin{aligned} b_X^{S_1,S_2} &= (b_x^{S_1} u_x^{S_2} + b_x^{S_2} u_x^{S_1})/\kappa \\ d_X^{S_1,S_2} &= (d_x^{S_1} u_x^{S_2} + d_x^{S_2} u_x^{S_1})/\kappa \\ u_X^{S_1,S_2} &= (u_x^{S_1} u_x^{S_2})/\kappa \\ a_X^{S_1,S_2} &= \frac{a_x^{S_1} u_x^{S_2} + a_x^{S_2} u_x^{S_1} - (a_x^{S_1} + a_x^{S_2}) u_x^{S_1} u_x^{S_2}}{u_x^{S_1} + u_x^{S_2} - 2u_x^{S_1} u_x^{S_2}}. \end{aligned} \tag{5.13}$$

where $\kappa = u_X^{S_1} + u_X^{S_2} - u_X^{S_1} u_X^{S_2}$. The assumption in the above formulation is that $\kappa \neq 0$.

5.1.5 Logical Operators and Opinions

In evidence theory the set of propositions is the power set of the chosen frame. The assumption was however that the power set is defined on a set consisting of only disjoint elements. In other words, only one elementary state of the system can be true at any given time. If for example, let us consider a set Θ consisting of disjoint elements $\{\theta_1, \theta_2, \theta_3\}$ such that $\theta_i \cap \theta_j = \phi, i \neq j$. The power set 2^{Θ} would contain the following elements $\{\theta_1, \theta_2, \theta_3, \{\theta_1, \theta_2\}, \{\theta_2, \theta_3\}, \{\theta_3, \theta_1\}, \Theta\}$. Standard set theory says the element $\{\theta_i, \theta_j\} = \theta_i \cup \theta_j$ implies θ_i or θ_j or both. However, if the assumption is that the elements $\theta_i, i = 1, 2, 3$ are disjoint, then the terms $\{\theta_i, \theta_j\}$ in the power set only means θ_i or θ_j. The case of set theoretic intersection of $\theta_i \cap \theta_j$ is not included. Hence if a belief mass assignment is designed with such an assumption on the frame then the probability expectation of the cases like $\theta_i \cup \theta_j$ or $\theta_i \cap \theta_j$ with $\theta_i \cap \theta_j \neq \phi, i \neq j$ cannot be computed using the definition in (5.7). Josang in [61] proposed a framework for applying logical operations like AND, OR and NOT on opinion tuples and probability expectations. The expressions for applying AND and OR operations on probability expectations are provided below.

Let Θ_1 and Θ_2 be two distinct binary frames of discernment with $X \in \Theta_1$ and $Y \in \Theta_2$ as two propositions. If the probability expectations of the propositions X and Y defined over Θ_1 and Θ_2 are $Pe(X)$ and $Pe(Y)$ respectively, then the combined

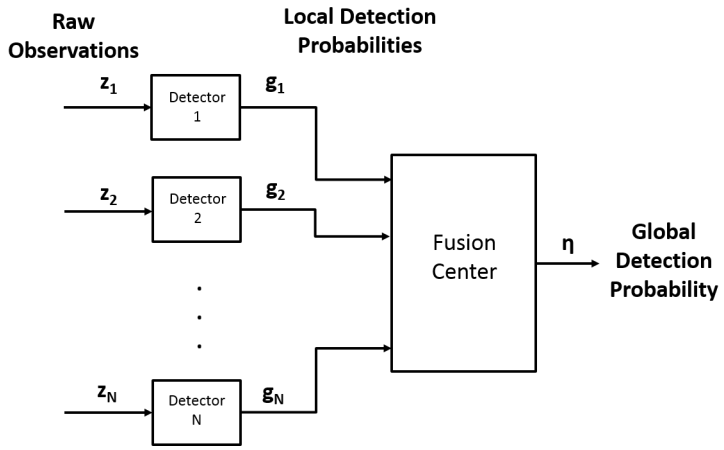


Figure 5.1: Distributed fusion of detection probabilities

probability expectations of $X \cap Y$ (logical AND) and $X \cup Y$ (logical OR) satisfy

$$Pe(X \cap Y) = Pe(X)Pe(Y) \tag{5.14}$$

$$Pe(X \cup Y) = Pe(X) + Pe(Y) - Pe(X)Pe(Y). \tag{5.15}$$

5.2 Fusion of Detection Probabilities

In the familiar scheme of a distributed detection system (see Figure 2.2), the goal at the fusion center is to combine local binary decisions to decide among a set of hypotheses. Contrary to this distributed decision fusion model, the framework developed by [26] assumes the same topology but allows the local detectors use the Bayes' rule locally to compute posterior detection probabilities which are then sent forward to the fusion center. The fusion center again uses Bayes' rule to combine the local detection probabilities to generate a global posterior detection probability. The fusion of probability design is shown in Figure 5.1. The scheme assumes a binary hypotheses set. We have also derived the generalized detection probability fusion rule

for multi-hypothesis set (see Appendix A).

In a system of N sensors observing a fixed volume represented by the binary hypothesis set $\Theta = \{\theta_1, \theta_2\}$, each sensor outputs the local detection probability of the occurrence of θ_1 based on its own observations z_i, $i = 1, ..., N$. We denote the local detection probabilities as $P(\theta_1|z_i) = g_i$. At the fusion center, $f(g_1, ..., g_N|\theta_1)$ is the conditional joint density of detection probabilities. The posterior detection probability at the fusion center can be defined as

$$\eta(g_1, ..., g_m) \quad = \quad P(\theta_1|g_1, ..., g_m). \tag{5.16}$$

Using Bayes rule and theorem of total probability the expression for posterior detection probability can be written as

$$\eta(g_1, ..., g_m) \quad = \quad \left(1 + \left[\frac{P(H_1)}{1 - P(H_0)} \frac{f(g_1, ..., g_N|\theta_1)}{f(g_1, ..., g_N|\theta_2)}\right]^{-1}\right)^{-1} \tag{5.17}$$

where $P(H_1)$ and $P(H_0)$ are the the prior probabilities of propositions θ_1 and θ_2 respectively. Equation (5.17) is the fusion rule to combine detection probabilities obtained from multiple sensors. Instead of detection probability, we choose to use the term *'degree of support'* since data obtained from sensors and also belief functions may not be exactly probabilities but rather probabilistic measures. The joint probability density functions $f(.)$ provides a statistical characterization of the sensor system and is determined by the individual design. Among the known parametric families of densities, the Beta family can be used to model the sensor data because they are rich

in shapes and most importantly their domain is the interval $[0, 1]$. Beta distribution is also the conjugate prior for Binomial events. Therefore, for a vast category of sensors, the Beta distribution can be used to model the likelihood functions. In the discussion that follows, the symmetric Beta distribution with different parameter sets (ν,ρ) are used to characterize the likelihood functions of the different sensors (the choice of symmetric beta distribution was merely to ease the mathematical calculations). The observations of each sensor are assumed to be independent of each other conditioned on the hypotheses.

5.3 Hard/Soft Fusion Algorithm

Once a finite set Θ of possible states forming the frame is decided, observers (human sources) are asked to assign belief masses or measures of confidence to the various propositions X which are members of 2^{Θ}. The human observer statements considered here are uncertain and represented as for example *I am 40% sure about proposition A and unsure about the others*. Such a statement can be modeled with a belief mass assignment toward a frame Θ as $m(A) = 0.4$, $m(\Theta) = 0.6$ and $m(X) = 0, \forall X \neq A, \Theta, X \subseteq \Theta$.

We assume $n = |2^{\Theta}| - 1$. The observers can base their decisions on prior knowledge, experience and/or data gathered or reported by other individuals, newspaper articles, television images etc. The belief mass assignments allow the computation of the opinion tuples and the corresponding probability expectations. Equations (5.14) and (5.15) must be used when the requirement is to obtain probability expectation values for union or intersection of elementary propositions. The probability expecta-

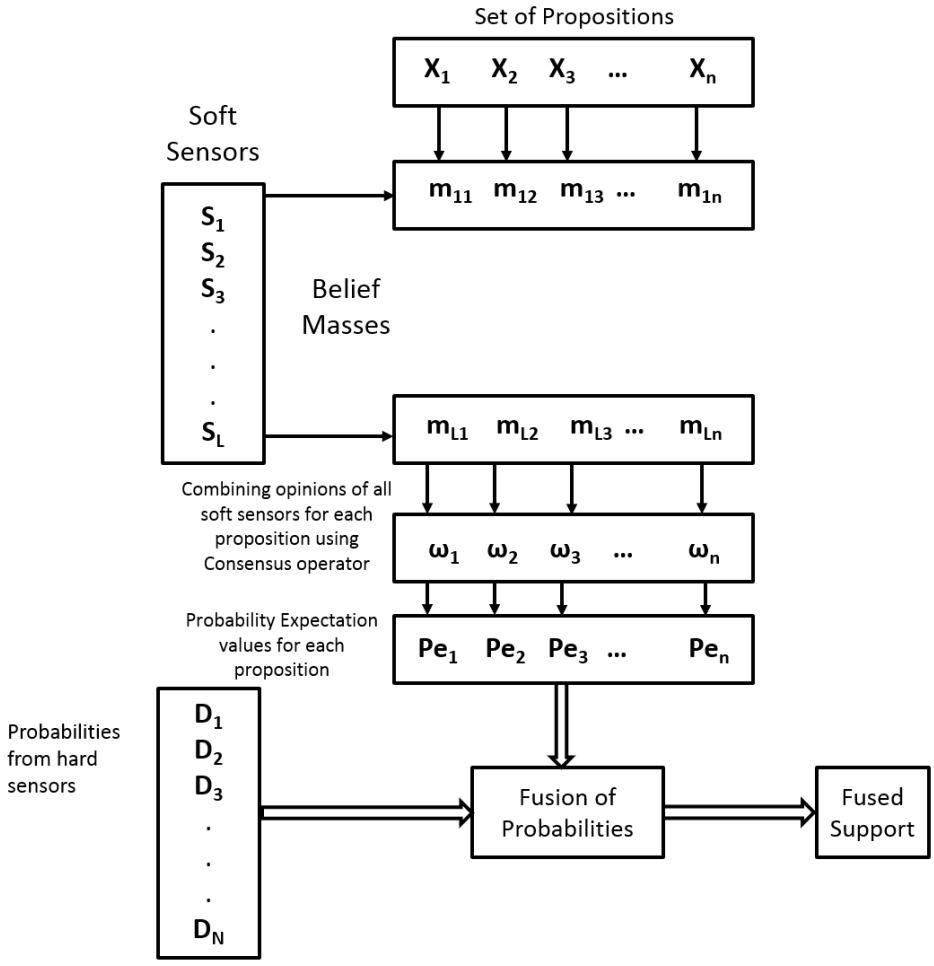


Figure 5.2: Framework for hard/soft fusion using evidence combination.

tion values can then be combined with the probabilistic measures obtained from hard sensors using a fusion of probability rule such as the one proposed by Krzysztofowicz and Long [26]. The process is shown in Figure 5.2. There are L soft sensors S_1 to S_L. All of them assign belief masses to each element of the proposition set, X_1 to X_n. In Figure 5.2, the assigned belief masses are represented by m_{ij} where i represents the soft sensor and j is the proposition number. The 4 coordinates needed to produce an opinion tuple (5.1) are computed and the corresponding opinions toward various propositions are formed. Each such opinion has an owner in the form of the soft sensor

who assigns belief masses to the proposition set. Each of the soft sensors S_1 to S_L will own opinions about each of the propositions in the set X_1 to X_n. The opinion held by soft sensor S_i about proposition X would be represented as $\omega_X^{S_i} = (b_X^{S_i}, d_X^{S_i}, u_X^{S_i}, a_X^{S_i})$. In the situation when multiple soft sensors provide opinions about a set of propositions, the Consensus operator is used to combine the various opinion tuples owned by different soft sensors but pertaining to the same proposition, to form the aggregate opinion tuple vector for that particular proposition. For the n propositions X_1 to X_n, these aggregate opinion tuples are represented by ω_1 to ω_n in Figure 5.2, each having a form as shown in (5.1). The consensus operator is associative and hence belief combination can be updated sequentially. However, if $n > 2$ beliefs are available simultaneously, then the following generalized consensus operator for combining opinions can be used. For opinion tuples $\omega_X^i = (b_X^i, d_X^i, u_X^i, a_X^i), i = 1, 2, ..., L$ held by L soft sensors about the same proposition X, the tuple $\omega_X^{1..L} = (b_X^{1..L}, d_X^{1..L}, u_X^{1..L}, a_X^{1..L})$ defined as

$$
\begin{aligned}
b_X^{1...L} &= \frac{\sum_{i=1}^{L} b_X^i \left(\prod_{j\neq i} u_X^j\right)}{\sum_{i=1}^{L}\left(\prod_{j\neq i} u_X^j\right) - (L-1)\left(\prod_{i=1}^{L} u_X^i\right)} \\
d_X^{1...L} &= \frac{\sum_{i=1}^{L} d_X^i \left(\prod_{j\neq i} u_X^j\right)}{\sum_{i=1}^{L}\left(\prod_{j\neq i} u_X^j\right) - (L-1)\left(\prod_{i=1}^{L} u_X^i\right)} \\
u_X^{1...L} &= \frac{\prod_{i=1}^{L} u_X^j}{\sum_{i=1}^{L}\left(\prod_{j\neq i} u_X^j\right) - (n-1)\left(\prod_{i=1}^{L} u_X^i\right)} \\
a_X^{1...L} &= \frac{\sum_{i=1}^{L}\left[a_X^i \left(\prod_{j\neq i} u_X^j\right)(1-u_L^i)\right]}{\sum_{i=1}^{L}\left(\prod_{j\neq i} u_X^j\right) - L\left(\prod_{i=1}^{L} u_X^i\right)},
\end{aligned}
\tag{5.18}
$$

represents the combined opinion tuple about proposition X. In the last expression of (5.18), a_X^i represents the focused relative atomicities obtained from (5.12). The combined values of belief $b_X^{1...L}$, uncertainty $u_X^{1...L}$ and relative atomicity $a_X^{1...L}$ are then used in (5.7) to provide the aggregate probability expectation value for the proposition X. Since the probability expectation of both the original and focused frames are the same, while using (5.7), the power set of the focused frame is used, i.e., containing sets X and $\overline{X}$ only. Hence, the expression for the aggregate probability expectation for a proposition X simplifies to

$$Pe(X)^{1...L} = b_X^{1...L} + (d_X^{1...L} \times 0) + (u_X^{1...L} \times a_X^{1...L}) \tag{5.19}$$

In Figure 5.2, these are represented as Pe_1 to Pe_n.

The hard sensor data are processed to produce a support values corresponding to each proposition. In other words, each hard sensor data will be a vector of support values each element of which is the support value corresponding to a particular proposition. The vector $D = [D_1, D_2, ..., D_n]$ in Figure 5.2 corresponds to the data from one hard sensor. The elements D_i, $i = 1, ..., n$ are the support values for the n propositions X_1 to X_n. In the scheme, there can be multiple such vectors each from a different hard sensor. All such vectors D from the hard sensors are fused with the probability expectation values obtained from (5.19), using the fusion rule described in (5.17) to generate fused support values for each proposition. A simple rule like choosing the proposition with maximum support can subsequently be used to make a decision about the frame.

While implementing the fusion rule in (5.17), there will be as many vectors of probabilistic support values as the total number of propositions. Each such vector will correspond to a particular proposition and will consist of support values obtained from all the sensors irrespective of hard and soft. Such a construction can be allowed since both the soft sensor and hard sensor data are now in the same format of probabilistic measures and there is no need to distinguish between them. The expression in (5.17) needs to be used for all such vectors separately to obtain fused support values for each member of the proposition set. Examples in the next section will elucidate this fact. We summarize the whole algorithm as follows:

1. Fix a set of cases to be considered and form the frame
2. Soft Sensors
    - Develop belief mass assignment toward the frame for each soft sensor.
    - Use (4.4), (5.2), (5.3) to form the partial focused opinion tuples (note the focused relative atomicities are yet to be computed).
    - Use (5.7) to form probability expectations for all propositions.
    - Use (5.12) to obtain focused relative atomicities.
    - Use (5.18) and (5.19) to form fused opinion tuples and probability expectations.
3. Hard Sensors
    - Process hard sensor data to obtain support values for all the propositions. This processing is application and data specific.

4. Use (5.17) to perform final fusion.

5.4 Examples

5.4.1 Example 1: Identity Recognition

This example simulates a situation with two soft sensors and two hard sensors and uses the data from each of them to calculate the final degree of support for the propositions under consideration.

Situation: A heist scenario was simulated where a valuable object was stolen from a secured place. Two persons A and B were asked to perform this act. The location was monitored using two cameras capable of sending live video feed to a computer. Let $\Theta = \{A, B, C\}$. The proposition "Unknown C" was added to Θ and caters to the case when the perpetrator is unidentified. This is a simulated situation and the truth is the set $(A \cup B)$.

Soft Sensor modeling: There can be the following two cases:

1. The elements of Θ are disjoint and the intersection $\theta_i \cap \theta_j = \phi, i \neq j$

2. The elements are not mutually exclusive and the intersection $\theta_i \cap \theta_j \neq \phi, i \neq j$

We do not provide an example supporting case 1 and perform the soft sensor modeling based on the assumption that any two propositions in Θ can occur simultaneously. The ternary set Θ was chosen as the frame. In addition to the probability expectations of the elements of Θ, we wish to find out probability expectations for the following cases as well: $(A \cup B)$, $(B \cup C)$, $(C \cup A)$ where the elements A, B and C are not disjoint i.e. $(A \cup B)$ means A or B or both and similar meaning holds for other unions

as well. Set theoretic operations are not applicable on opinion tuples and probability expectation calculations unless the frame of discernments are distinct [61]. In other words, equations (5.14) and (5.15) can be applied only when there are two distinct frames $\Theta_1 = \{X, \overline{X}\}$ and $\Theta_2 = \{Y, \overline{Y}\}$ with $X \cap Y = \phi$, $X \cap \overline{Y} = \phi$ and $\overline{X} \cap Y = \phi$. For this example we therefore define 3 distinct sets of propositions as follows:

$$F_1 = \{A, \overline{A}, \Theta_A = \{A, \overline{A}\}\};$$

$$F_2 = \{B, \overline{B}, \Theta_B = \{B, \overline{B}\}\}; \quad F_3 = \{C, \overline{C}, \Theta_C = \{C, \overline{C}\}\}.$$

To obtain the belief masses, video snapshots captured from the cameras were presented to 18 different individuals and all of them were asked to independently provide feedback in the form of confidence in the occurrence of the propositions. The individuals were asked to provide a number between 1 and 10, higher value representing more confidence. The data was collected independently for the 3 sets of propositions. The normalized set of those values were used as the belief mass assignments for the propositions. Depending upon the situation, the individuals can be witnesses, investigators or experts. Henceforth the data obtained from them will be referred to as just soft sensor data. Two soft sensors were randomly sampled from the set of 18 and corresponding belief masses were used for calculation of probability expectations. The sampling was repeated 100 times. Table 5.1 lists one such set of belief mass assignment for soft sensors S_1 and S_2.

An opinion will be represented as $\omega_X^{S_i} = (b_X^{S_i}, d_X^{S_i}, u_X^{S_i}, a_X^{S_i})$, $i = 1, 2$. S_i denotes the owner of the opinion and X is the proposition in focus. The parameters $(b_X^{S_i}, d_X^{S_i}, u_X^{S_i}, a_X^{S_i})$ are the belief, disbelief, uncertainty and focused relative atomicity

Table 5.1: Initial belief mass assignments from soft sensors for example 1.

Propositions	**Belief** b^{S_1}	**Belief** b^{S_2}
A	0.488	0.130
$\overline{A}$	0.366	0.779
Θ_A	0.024	0.065
B	0.727	0.501
$\overline{B}$	0.182	0.333
Θ_B	0.091	0.166
C	0.539	0.2
$\overline{C}$	0.385	0.6
Θ_C	0.077	0.2

respectively of the proposition X according to owner S_i. The focused opinion tuples for the belief mass assignments in Table 5.1 are presented below.

$$\omega_A^{S_1} = (0.488, 0.366, 0.024, 1/2); \quad \omega_A^{S_2} = (0.130, 0.779, 0.065, 1/2)$$

$$\omega_B^{S_1} = (0.727, 0.182, 0.091, 1/2); \quad \omega_B^{S_2} = (0.501, 0.333, 0.166, 1/2)$$

$$\omega_C^{S_1} = (0.539, 0.385, 0.077, 1/2); \quad \omega_C^{S_2} = (0.200, 0.600, 0.200, 1/2)$$

Note that the relative atomicity in the above opinions is 1/2 since the distinct sets of propositions are formed from binary frames of discernment. The combined opinion tuples and the probability expectations toward each proposition were calculated using (5.18) and (5.19). The aggregate probability expectations are provided below.

$$Pe^{S_1,S_2}(A) = 0.250, \quad Pe^{S_1,S_2}(B) = 0.442, \quad Pe^{S_1,S_2}(C) = 0.308.$$

Using the expression in (5.15) for applying logical OR operation to opinion tuples, the following probability expectation values were computed:

$$Pe^{S_1,S_2}(A \cup B) = 0.581, \quad Pe^{S_1,S_2}(B \cup C) = 0.614, \quad Pe^{S_1,S_2}(C \cup A) = 0.481.$$

The vector of probability expectations toward the propositions $\{A, B, C, (A \cup B), (B \cup C), (C \cup A)\}$ so obtained by averaging over 100 runs was used for fusion with hard sensor data.

Hard Sensor modeling: The hard sensors are 2 cameras which are situated at two different positions monitoring the location of the object and can provide video data. A database of 118 images was created. Image analysis was done by a face recognition software [79] which compared test images with those stored in the database and also provided corresponding degrees of similarity (based on the norm of the vector between average eigen faces of each image class and the test image). The images of the entities A and B which were well defined propositions in the soft sensor model were included in the database and were tagged (the scheme can be shown to work even in the case where images of entities included as members in the frame are not recorded). Image frames were extracted from the video captured by the cameras and a face extraction algorithm was used to separate the face regions in the image frame. These sub images of faces were suitably resized and were compared individually with the database. A threshold was chosen with the goal that any degree of similarity above that threshold would be considered as a match. The procedure to obtain the degree of support values

is explained below. Taking into account that the prime suspects are A and B, the following cases were considered.

Case 1: Let the maximum cardinality of elements of the proposition set be $card_{max}$. In this example $card_{max} = 2$. Lets consider the case when the image frame contains $f \geq card_{max}$ face images, two of which separately have their maximum degrees of similarity with A and B whereas the other $(f-2)$ images do not have a match with anyone in the database. In such a situation, the maximum degrees of similarity with A and B are taken to be support values for the corresponding propositions A, B respectively and the $(f-2)$ faces are ignored since there is no way to compare them against any reference. On the other hand if k images ($k \leq f-2$), have matches with the database (i.e., k images when compared individually against the database, provide k corresponding maximum degrees of similarity all of them being above the threshold), the average of the k maximum degrees of similarity is considered to be the support value for the proposition C. The support values for any composite propositions are obtained as the sum of the support values of the constituent propositions. For example, the support value for the proposition $(A \cup C)$ would be the sum of the supports for A and C.

Case 2: When the frame contains only 2 facial regions and both of them match with A and B, then the support for the proposition C is taken to be a very small number $\epsilon > 0$ and the support for the non atomic propositions containing C is calculated as the sum of ϵ and the support value of the other constituent proposition. For example, support for $(A \cup C)$ would be *support of* $A + \epsilon$.

Case 3: When k faces match with images in the database, none of them being either

A and B, then the average of the k maximum degrees of similarity is taken to be the support for proposition C. The support values for propositions A and B are taken to be ϵ.

Case 4: If the number of facial regions detected in the frame is $f < card_{max}$, then the support values for the propositions with cardinality greater than f are taken to be a small number $\epsilon > 0$.

Case 5: When among k facial regions which match with entries in the database, only one matches with either of the propositions A and B, then average of the remaining $k-1$ maximum degrees of similarity is considered to be the support for the proposition C. The support for the proposition among A and B with which there was no match is assigned a small number $\epsilon > 0$.

For the subsequent steps we chose $\epsilon = 0.05$. In the simulation, one camera (hard sensor 2 in Table 5.3) feed had a single facial region and hence for this data, all propositions with more than 1 constituent element were assigned a support value of 0.05. For the other camera (hard sensor 3 in Table 5.3), the video had two facial regions which matched with A and B. This above situation resulted in non conformity between camera results since for one of the camera feeds, only one face was detected and therefore the corresponding decision supported a singleton proposition whereas the other camera data with two faces, supported one of the non atomic cases. This closely resembles a situation in reality where due to reasons like view angle, lighting conditions, occlusion etc., cameras at different vantage points do not see the same objects in similar number or state.

Without loss of generality in this example scenario, the soft sensor data which

Table 5.2: Sensor parameters.

Hypothesis	**Sensor 1(Soft)**	**Sensor 2(Hard)**	**Sensor 2(Hard)**
h=1(alternate)	ν=3,ρ=2	ν=3,ρ=1	ν=4,ρ=1
h=0(null)	ν=2,ρ=3	ν=1,ρ=3	ν=1,ρ=4

has been referred to in terms of probability expectation, was assumed to be beta distributed with certain set of parameters. The soft sensor thereby was included as a part of the cluster of 3 sensors. All of them produced an output vector of degree of support (each being a 6 element vector for 6 propositions). The corresponding likelihood functions were modeled using symmetric Beta distribution with each sensor having a different set of (ν,ρ) parameters under different hypotheses. The hard sensor (essentially the face recognition algorithm associated with each camera) Beta distribution parameters were empirically estimated with $\nu/(\nu+\rho)$ representing frequency of correct identifications and $\rho/(\nu+\rho)$ representing the frequency of incorrect identifications when the face recognition algorithms were tested against a training set.

Recall that belief combination is performed using Consensus operator by using focused binary frames essentially containing two hypotheses. Therefore for each focused frame, $h = 1$ in Table 5.2 represents the proposition in focus is true whereas $h = 0$ represents the complement of the focused proposition is true.

Two soft sensors were randomly sampled from a group of 18. Their opinions were combined and a probability expectation vector obtained. The sampling was done 100 times and an averaged probability expectation vector was computed. The two

Table 5.3: Unnormalized support assignments for example 1.

Sensors	**Propositions**					
	A	B	C	$A \cup B$	$B \cup C$	$C \cup A$
1	0.268	0.326	0.227	0.773	0.732	0.674
2	0.234	0.288	0.237	0.05	0.05	0.05
3	0.305	0.274	0.05	0.579	0.279	0.310

Table 5.4: Normalized combined support values toward each proposition for example 1.

	Propositions					
	A	B	C	$A \cup B$	$B \cup C$	$C \cup A$
Final Fusion	0.090	0.132	0.001	0.748	0.014	0.016

camera data vectors were then fused with this average probability expectation vector. Table 5.3 shows the support values (unnormalized probability expectation values from soft sensors averaged over 100 runs and degree of support values from hard sensors) for the various propositions.

Each column in this table forms a vector $g = [g_1 \ g_2 \ g_3]$ and the elements g_i, $i = 1, 2, 3$, are the support values from all sensors toward a particular proposition. Using (5.17) for each column in Table 5.3, the final support values for the propositions were calculated. Table 5.4 shows the final normalized result against the propositions.

It can be inferred that the opinions from human experts and hard sensor data when combined, results in a cumulative value which supports the proposition $(A \cup B)$ to the maximum extent. This coincides with the actual truth.

In almost all real scenarios the truth is unknown and the proposition which the

above technique will support depends heavily on the belief mass assignments and the efficiency of the hard sensor data processing software used. The above process can be employed in an expert system followed by a decision making module. Depending upon the application, a decision rule like a threshold comparison can be used to produce an informed result.

5.4.2 Example 2: Object Localization

The example simulates a situation with four soft sensors and one hard sensor and uses the data from each of them to calculate the final degree of support for the propositions under consideration.

Situation: Let there be a possible geographic region which is of interest for undertaking archaeological digging. This location was chosen near Valley Forge Memorial Park, Philadelphia. A feasible latitude-longitude position was fixed and assumed to be the true location. Four persons were asked to find out the location of the site within a region of 0.1 mile radius. None of them had any knowledge of the true position. One of the persons had a Global Positioning System (GPS) tracker attached to him which could send out position data to a centrally located server. All the individuals were also given pictures of how the location would roughly look like physically. It was also assumed that the individuals never got separated during their search.

Two cases were considered: the persons are within 0.1 mile of the site denoted by I, the persons are outside 0.1 mile region denoted by O. The humans could see the geographical location and compare with the pictures. This helped the observers to send out reports with answers to the 2 propositions under consideration. A similar

Table 5.5: Belief support assignments for example 2.

Propositions	**Sensor1 (Soft)**	**Sensor2 (Soft)**	**Sensor3 (Soft)**	**Sensor4 (Soft)**	**Sensor5 (Hard)**
<0.1 mile(I)	0.5	0	0.1	0	1
>0.1 mile(O)	0.4	0.9	0.8	0.99	0.01
Uncertainty	0.1	0.1	0.1	0.01	-

feedback scheme as in example 1 was used. The tracker provided just the hard location. The location from the tracker data was compared using Google Earth and support values for the two propositions were obtained. This data was regarded as the hard sensor data. The process is described as follows: Using the location (latitude-longitude) from the tracker, its line of sight distance from the assumed true location was computed. If the calculated distance is less than 0.1 miles, the first proposition I was assigned a value of 1 and the other O a small value of 0.01 and vice versa. The small value of 0.01 was assigned instead of zero so that no undefined operations like division by zero were encountered. Table 5.5 shows a set of support values from the persons in the field (soft sensors) and the GPS tracker (hard sensor) when the whole set up was simulated.

The uncertainty of the GPS tracker can be represented in terms of percentage of correctness of its output. This work did not use any value for that proposition corresponding to the hard sensor. Some of the soft sensors were in disagreement as can be noted from Table 5.5. The Consensus operator was used to combine the conflicting opinions and the probability fusion rule allowed to aggregate the hard sensor data and the combined opinion to obtain a fused result. The final support

Table 5.6: Combined support values for each proposition for example 2.

	Propositions	
	I	O
Final Fusion	0.841	0.159

values for the various propositions as calculated are provided in Table 5.6. The time when the data in Table 5.5 was obtained, the group was actually within 0.1 mile of the true location. The final fused result in Table 5.6 conforms that.

5.4.3 Example 3: Image Analysis

This section illustrates a scenario where a set of satellite images of a particular geographical location and a set of text based data sources are collected and fused. Figure 5.3 shows satellite imagery of a couple of newly constructed sites in a location called Semnan in Iran. These were captured by the Eros-B satellite in 2009 (the images are widely available on the Internet). There are many articles in newspapers, defense magazines and on line discussion forums speculating the possible reasons for such massive construction. Barring the speculations, the unanimous opinion has been that these constructions are part of a large missile base. This example shows how the hard satellite image data and soft text based information can be combined to get support values for a possible set of suitable propositions.

As a first step toward hard sensor data processing, the suitable satellite images were selected and passed as inputs to an edge extraction module. The images where edge extraction is very difficult or there are too many objects were not used for edge analysis. The edge extraction module processed the image in two steps. First

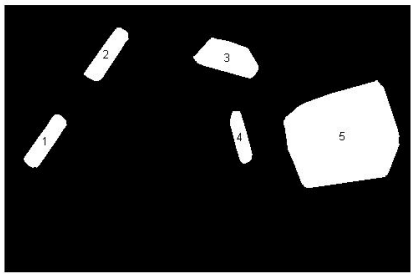
(a) Possible missile assembly location.

(b) Possible missile launch Site.

Figure 5.3: Eros-B satellite images showing newly constructed sites.

it removed the shadows (if any) and second, it filtered the image to extract closed regions. In this example, the image in Figure. 5.3a was used. Figure 5.4 shows the result of this processing using a generic shadow removal algorithm based on pixel intensity comparison and the standard Sobel filter [80] for edge extraction. Using an a priori specified image scale (meter to number of pixel ratio), the areas and perimeters of the buildings were obtained. These data points would be helpful in understanding whether the buildings are big enough to house missile assembly units or manufacturing equipments. The aforementioned information and the images were provided to a domain expert for analysis and the support values to the set of propositions were asked. A collection of text based sources each disjoint from each other in the sense that neither of them contain same information or are written by the same author were grouped as the soft data. Two human interpreters were asked to read through this data set and find threads of importance and relevance considering the proposition set. The human interpreters provided a set of support values toward

(a) Image from Figure. 5.3a with shadows removed.



(b) Closed regions resembling buildings.

Figure 5.4: Image processing (Shadow removal and edge extraction) of satellite images.

the propositions under consideration, possibly based on abstract and complex traits like experience, missing clues, information both historical and political.

The set of propositions considered in this example are listed below.

- A-The missile base is a new development to cater to a new class of long range missiles.
- B-The construction is a sister facility to a larger, already existing base and poses no new threat.
- C-The construction is shifting of an existing base and abandonment of the latter and is purely for logistical and economic reasons.

Table 5.7 shows the support values for the set of propositions under consideration. The fusion of the soft data was performed using Consensus operator and the final fusion with hard sensor data was performed using (5.17). Table 5.8 lists the final

Table 5.7: Support assignments for example 3.

Propositions	**Sensor1 (Hard)**	**Sensor2 (Soft)**	**Sensor3 (Soft)**
A	0.5	0.8	0.2
B	0.1	0	0.2
C	0.3	0.1	0.4
Uncertainty	0.1	0.1	0.2

Table 5.8: Combined support values for each proposition for example 3.

	Propositions		
	A	B	C
Final Fusion	0.727	0.022	0.250

fused support values. No fusion scheme is universal and based on the propositions, the processing module that analyzes data becomes a dynamic function. The processing done for the above example may not be useful for generating support values toward a different proposition set. However, the example shows that disparate and heterogeneous sources can be combined to yield a result that can facilitate informed decision making.

6: HIERARCHICAL EVIDENCE TREES

The belief theory framework was originally introduced by Shafer in [16] as a method of modeling uncertain evidence. The assignment of belief masses to sets of propositions and building a support function over a power set handle ignorance and uncertainty of sources in an intuitive way. Such a modeling method seems very apt to be used to model opinions coming in from human observers as such data lacks the well defined structure required for the use of traditional probabilistic reasoning.

For example, suppose that Θ represents a set of mutually exclusive propositions, then in belief theory, an opinion of an observer who is 40% sure of a particular proposition $A \subseteq \Theta$ being true and is ignorant about the other possibilities, can be represented by a belief mass assignment $m : 2^{\Theta} \mapsto [0, 1]$ with $m(A) = 0.4$, $m(\Theta) = 0.6$ and $m(X) = 0, \forall X \neq A, \Theta$. However, in many practical applications of belief functions, not all opinions are straightforward and might pertain to multiple questions, answers to which are related in a hierarchical manner. Each question has multiple possible answers with only one of them being actually true. The set of possible answers to each question form an exhaustive set of mutually exclusive elements essentially a frame of discernment. Such hierarchically related frames can be modeled using tree structures.

Work on hierarchical frames was initiated by Gordon and Shortliffe in [81]. They were concerned with the problem of combining evidence using Dempster's rule of Combination when different items of evidence were associated to frames of varied

refinement. The objective was to implement Dempster's rule in a computationally tractable form and in the process Gordon and Shortliffe developed an approximation for the combination rule. Shafer and Logan [82] took Gordon and Shortliffe's work forward and proposed a method to exactly implement Dempster's rule in a manner such that the computational complexity is linear. Pearl in [83] developed a Bayesian approach to the problem of combining evidence in a hierarchical hypothesis space. In [84], a work on propagation of Bayesian probability judgments in causal trees was proposed. Some of the other prominent works pertaining to belief propagation in trees have been [85], [86] and [87]. In all of these works the assumption has been that there is a fixed root frame, subsets of which branch out according to some partition and form the various nodes of the tree. In Pearl's work, the nodes of a tree correspond to random variables and directions of the links are interpreted as directions of causation. Thus each variable is influenced by variables above it in the tree and influences the variables below it.

In the following discussion, we retain the idea of causality but generate the tree with the assumption that each node is not a subset of a fixed root frame but rather a different frame in itself. We introduce a class of hierarchically related opinions termed as *Conditionally Refined opinion (CRo)* and discuss their formulation. CRo would describe the opinion which a human observer S would most likely provide under the following circumstance:

S has some confidence toward a set of disjoint hypotheses. Also, conditioned on each such hypothesis and independent of the confidence toward it, W may hold an opinion toward a proposition belonging to a refinement of the hypothesis. For instance, an

opinion of the form *"I am 60% sure about the target being an aircraft and 40% sure about it being a missile but in case its an aircraft, there is high chance it is a fighter."* would be a CRo.

We assume a class-subclass relationship between the elements of various frames, i.e a child frame is a more refined representation of its parent element and a child frame would contain distinct elements of its own which are not contained exactly in its parent frame. The child frame can be caused only if the parent element is plausible. A frame is a child frame in the sense that its elements represent special cases of the broader class which is its parent. Also in the works presented in [81, 82, 87], the assumption was made that the belief functions assigned to various frames are simple support or dichotomous functions. This work holds no such assumptions and can be applied for any form of belief assignments. We also develop algorithms for belief mass propagation down the tree. The proposed representation is shown to be applicable to various soft-soft and hard-soft fusion situations. Using the advantages of the organization of the tree, all belief combination calculations were performed using small frames and later combined together by a simple concatenation operation making the proposed scheme a computationally attractive framework.

6.1 Generalized Hierarchical Evidence Structure

Let there be a CRo provided by some observer where the support values are provided toward a set of hypotheses denoted by Ω and also on its various subclasses or refinements. In this context, the goal is to create a multi-level structure which maintains the hierarchical relationships that exist between frames and also allows fusion.

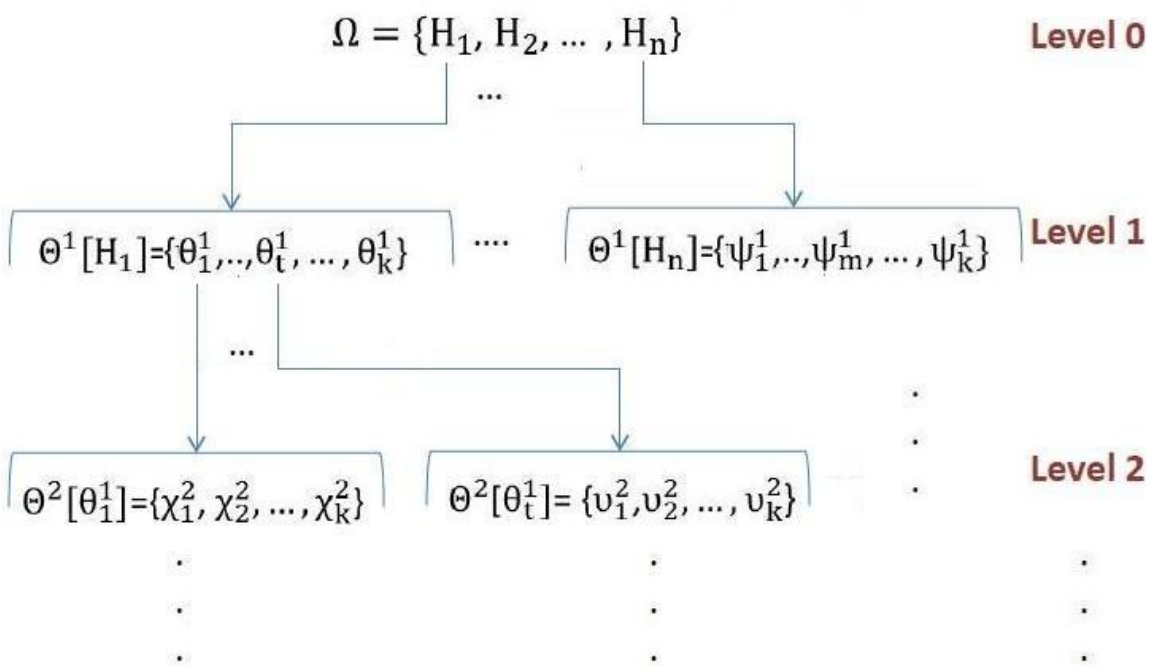


Figure 6.1: Generalized hierarchical structure of belief frames.

Figure 6.1 depicts the general multi-level organization. Each element in a frame is assumed to represent a class and can be divided into subclasses. In that case, the element being divided would be a parent element and the set of subclasses would form the child frame. This structure is similar to a Bayes' network [88] construction, however here the nodes are frames of mutually disjoint elements instead of random variables. Furthermore, there are in general multiple elements in a node (elements of the frame) with each element allowed to have its respective child frames. The hierarchy of Figure 6.1 is a tree in the strict sense - each child frame below Level 0 has an unique parent. We refer to all elements without any child frame as *leaf elements*. All leaf elements denote the end of the path connecting itself and the ancestor in the root frame. In the subsequent discussion we will denote the collection of leaf

elements belonging to frames in levels up to and including level $j-1$ as R^{j-1} and an arbitrary element of this set would be represented as r. The frame Ω in Level 0 is referred as the root node or root frame and represents the set of mutually exclusive hypotheses in the coarsest sense. Each hypothesis H_i represents a superclass and can be refined into a set of subclasses. All such sets would then contain disjoint elements each representing specific subclasses and therefore would form the child frames. In the following discussion, we would represent all child frames (frames in Level 1 and downward) with the notation $\Theta^j[P]$, where Θ is the common specifier for a child frame and P within square brackets indicate the parent element and j is the level. For example in Figure 6.1, $\Theta^1[H_1]$ at level 1 represents the child frame of hypothesis H_1 and $\Theta^2[\theta_1^1]$ at level 2 is the child frame of hypothesis θ_1^1 which in turn is an element of the child frame $\Theta^1[H_1]$. By construction a frame is disjoint with any element in its ancestor level which is not its parent and also with any frame in the same level since each frame has unique and disjoint parents. For this generalized structure, we provide the formulation for belief propagation in the following sections.

6.1.1 Propagating Belief Masses of Singleton Elements of a Parent Frame

Every frame in the tree has an associated belief mass function and all child frames emanate from distinct elements of its parent frame. We discuss in this section how to propagate the belief masses of the singleton parent elements to their child frames. For the sake of explanation we will use the 3 level tree shown in Figure 6.2 as a running example. In the generalized (M level tree) context, for each child frame from level

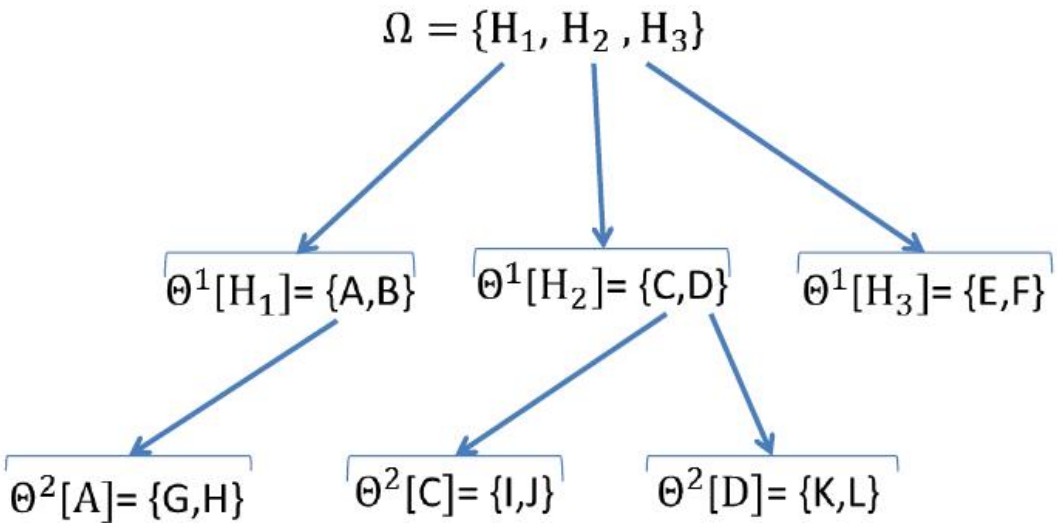


Figure 6.2: A three level evidence tree.

j, $j > 0$ onwards, the observer provides a normalized belief mass assignment. Let us consider a child frame $\Theta^j[P]$ in level j, $j \in \{1, .., M\}$ (we assume there are M levels of refinement) with P as its parent element which by construction must be an element of a frame in level $j-1$. Let P be an element of the frame Q. If the normalized belief mass assigned toward $\Theta^j[P]$ is represented as $m'_{\Theta^j[P]}$ then it must satisfy

$$\sum_{X \subseteq \Theta^j[P]} m'_{\Theta^j[P]}(X) = 1. \tag{6.1}$$

However, for the opinions in the class of CRo, the observer assigns these masses only assuming P is true. Hence we consider the masses $m'_{\Theta^j[P]}$ as conditional masses and denote them as $m_{\Theta^j[P]}(X|P)$. Then we define the posteriori belief masses $m_{\Theta^j[P]}$, $\forall X \subseteq \Theta^j[P], X \neq \phi$ as

$$m_{\Theta^j[P]}(X) = m_{\Theta^j[P]}(X|P) m_Q(P), \tag{6.2}$$

where $m_Q(P)$ is the belief mass of the element P in frame Q. We consider $m_Q(P)$ to be the total evidence available to be distributed (if required) between subsets of the refinement $\Theta^j[P]$ and therefore the scaled belief mass $m_{\Theta^j[P]}$ is computed such that $m_{\Theta^j[P]} : 2^{\Theta^j[P]} \mapsto [0,1]$ and

$$\sum_{X \subseteq \Theta^j[P]} m_{\Theta^j[P]}(X) = m_Q(P). \tag{6.3}$$

The above scaling is performed for all normalized belief mass assignments of child frames at all levels except the root level. For any child frame at level $j > 1$, the scaled belief mass of the parent element in level $j-1$ should be used to perform (6.2). In the context of the example tree in Figure 6.2, for simplicity sake consider the case when the belief mass assignment of the root frame is Bayesian with belief masses given as $m_\Omega(H_1) = m_1$, $m_\Omega(H_2) = m_2$, $m_\Omega(H_3) = m_3$ and $m_\Omega(\Omega) = 0$. The belief masses assigned on Ω satisfy $\sum_i m_i = 1$. The whole process of belief propagation is a top-down approach which starts at level 1 and proceeds downwards along the tree. Therefore, starting with the frames at level 1, the conditional belief masses of the child frames are scaled as

$$\begin{aligned}
m_{\Theta^1[H_1]}(X) &= m_{\Theta^1[H_1]}(X|H_1)p_1, \forall X \subseteq \Theta^1[H_1] \\
m_{\Theta^1[H_2]}(X) &= m_{\Theta^1[H_2]}(X|H_2)p_2, \forall X \subseteq \Theta^1[H_2] \\
m_{\Theta^1[H_3]}(X) &= m_{\Theta^1[H_3]}(X|H_3)p_3, \forall X \subseteq \Theta^1[H_3].
\end{aligned}$$

The belief masses of the child frames in level 2 are scaled as follows

$$
\begin{aligned}
m_{\Theta^2[A]}(X) &= m_{\Theta^2[A]}(X|A)m_{\Theta^1[H_1]}(A), \forall X \subseteq \Theta^2[A] \\
m_{\Theta^2[C]}(X) &= m_{\Theta^2[C]}(X|C)m_{\Theta^1[H_2]}(C), \forall X \subseteq \Theta^2[C] \\
m_{\Theta^2[D]}(X) &= m_{\Theta^2[D]}(X|D)m_{\Theta^1[H_2]}(D), \forall X \subseteq \Theta^2[D].
\end{aligned}
$$

6.1.2 Propagating Belief Masses of Composite Subsets of a Parent Frame

In a generalized M level tree, let us consider the frame $Q = \{Q_1, ..., Q_k\}$ at level $j-1$. A belief mass assigned toward this frame will be normalized in case Q is the root frame or else it will be scaled by the mass of its parent element as discussed in the preceding section. All the singleton elements Q_i are disjoint and can be parents to different child frames. In the last section we discussed how to transfer the belief masses of the singletons Q_i to subsets of their respective child frames. However, after this distribution the belief masses assigned to composite subsets of Q must also be taken into consideration. The aim is not to increase uncertainty but to maintain its value as one goes down the tree. Therefore, if $m_Q(X)$ is a nonzero belief mass toward X, $X \subseteq Q$ with $X = \{Q_{i_1}, ..., Q_{i_l}\}, i_t \in \{1, ..., k\}, l \leq k$ and $\Theta^j[Q_{i_t}]$, $t = 1...l$ are the child frames of the elements in X, then the mass $m_Q(X)$ is assigned to the set $W = \{\Theta^j[Q_{i_1}] \cup ... \cup \Theta^j[Q_{i_l}]\}$. For all Q_{i_t} which do not have any child frame (in other words the leaf elements of level $j-1$), $\Theta^j[Q_{i_t}]$ is replaced by Q_{i_t} in W. Furthermore, defined over every level j, let V^j be the set of subsets like W. In other words, V^j contains all the subsets in a particular level to which the belief masses of the composite

subsets of their respective parent frames are assigned. At any particular level j, let us assume there are k child frames $\Theta^j[P_i]$, $i = 1...k$ with P_i as their respective distinct parent elements. By taking the union of all frames in a particular level j with the elements of the set R^{j-1}, we can form a consolidated frame that encompasses the whole space of refinement i.e the entire gamut of information contained in the root frame Ω. We define such a frame at level j as

$$\Psi = (\bigcup_i \Theta^j[P_i]) \cup (\bigcup_{g \in R^{j-1}} g). \tag{6.4}$$

Ψ can be thought of as a general decision layer of the tree. It includes all members of the frames at level j, and all leafs occurring before level j. We do not associate a level identifier in the notation of Ψ since Ψ may contain elements from multiple levels. Note that members of the set V^j would be contained in the power set of the consolidated frame Ψ.

For illustration purposes, again going back to Figure 6.2, we note that since m_Ω is Bayesian, its focal set does not have any composite subsets of Ω. Therefore the set V^1 is empty. However the belief masses of the composite subsets of each frame in level 1 needs to be redistributed to subsets comprising elements of level 2. The leaf elements are present on level 1 and level 2 with $R^1 = \{B, E, F\}$ and $R^2 = \{G, H, B, I, J, K, L, E, F\}$. Traversing from level 1 to level 2, the scaled belief masses of composite elements of the frames $\Theta^1[H_1], \Theta^1[H_2]$ and $\Theta^1[H_3]$ are re assigned as

$$m_{\Theta^1[H_1]}(A \cup B) \rightarrow \{\Theta^2[A] \cup B\} \tag{6.5}$$

$$m_{\Theta^1[H_2]}(C \cup D) \rightarrow \{\Theta^2[C] \cup \Theta^2[D]\}, \tag{6.6}$$

where the symbol "$\rightarrow$" means "assigned to the set". Note that in (6.5) the element B is included since it is a leaf element and a member of R^1. As H_1 is the root ancestor of B and a part of the mass assigned to H_1 is contained in B, its exclusion in (6.5) would mean only partial information being propagated along the tree which is undesirable. The consolidated frames at the 3 levels of the tree are mentioned below:

$$Level0 : \Psi = \{H_1, H_2, H_3\} \tag{6.7}$$

$$Level1 : \Psi = \{A, B, C, D, E, F\} \tag{6.8}$$

$$Level2 : \Psi = \{G, H, B, I, J, K, L, E, F\}. \tag{6.9}$$

An attractive feature of the tree structure developed here is that multiple frames like Ψ can be produced based on requirements and all the elements of such frames need not belong to the same level. The couple of constraints are that the elements of the frame together must cover the entire information or hypotheses space defined by the root frame and there must not be any redundancy, i.e if a parent frame is included in Ψ, it's child frames should not be included as well. For any such frame Ψ, a pignistic probability vector can be obtained and used in subsequent decision making. As an example, considering the tree in Fig. 6.2, some possible consolidated frames might

take the following forms

$$\begin{aligned}
\Psi &= \{A, B, I, J, K, L, E, F\} \\
\Psi &= \{G, H, B, C, D, E, F\} \\
\Psi &= \{G, H, B, C, K, L, E, F\} \\
\Psi &= \{G, H, B, I, J, D, E, F\} \\
\Psi &= \{A, B, I, J, D, E, F\}
\end{aligned}$$

However a frame like $\Psi = \{G, H, I, J, K, L\}$ will not fit into the given structure since it contains no information about root hypothesis H_2 and offers only partial information about H_1 due to the exclusion of the element B. Therefore pignistic probabilities (4.25) for elements of this frame would not add up to unity.

Recall that the pignistic probability transform defined by Smets in [67] is given by

$$BetP(X) = \sum_{Y \subseteq \Delta, X \in Y} \frac{1}{|Y|} \frac{m(Y)}{(1 - m(\phi))}, \forall Y \in \Delta. \tag{6.10}$$

In this development, we use a non-normalized version of the pignistic transformation defined over any arbitrary frame Δ as

$$BetP^{\Delta}(X) = \sum_{Y \subseteq \Delta, X \in Y} \frac{m(Y)}{|Y|}, \forall X \in \Delta. \tag{6.11}$$

The transformation in (6.11) is applied to every frame on a particular level. Since the scaled belief masses of any frame satisfy (6.3), the pignistic transformation for each

child frame will satisfy

$$\sum_{X \in \Theta^j[P]} BetP^{\Theta^j[P]}(X) = m_Q(P), \tag{6.12}$$

where $\Theta^j[P]$ is the frame at level j on which the transformation is being applied and P is its parent element belonging to frame Q in level $j-1$. The reason for using a non normalized version of the pignistic transformation is so that (6.12) is satisfied. A normalized pignistic transformation would have resulted in the sum in (6.12) to be always unity. Let $BetP^j$ represent the vector of pignistic probabilities obtained by applying the transformation on the set V^j at level j and $BetP^{\Theta^j[P]}$ represent the pignistic probabilities obtained from the frame $\Theta^j[P]$. The dimensions of the vectors $BetP^{\Theta^j[P]}$ and $BetP^j$ are $|\Theta^j[P]| \times 1$ and $|\Psi| \times 1$ where $|\Theta^j[P]|$ and $|\Psi|$ are the sizes of the frames $\Theta^j[P]$ and Ψ respectively and Ψ is the consolidated frame at level j as defined by (6.4). For a certain element of Ψ which is not included in any of the subsets in V^j, its pignistic probability in the $BetP^j$ vector will be 0. If there are k child frames at level j with corresponding parent elements P_i, $i = 1...k$ then the following holds

$$|R^{j-1}| + \sum_{i=1}^{k} |\Theta^j[P_i]| = |\Psi|. \tag{6.13}$$

We form the vector Γ^{Rj-1} with each element $m(r)$ holding the scaled belief mass of a particular $r \in R^{j-1}$. Since the combined frame Ψ at level j contains elements of R^{j-1} along with the child frames $\Theta^j[P_i]$, $i = 1...k$, we form the pignistic probability vector

of Ψ at a specific level j as

$$Pr = \begin{bmatrix} BetP^{\Theta^j[P_1]} \\ \vdots \\ BetP^{\Theta^j[P_k]} \\ \Gamma^{R^{j-1}} \end{bmatrix} + \begin{bmatrix} BetP^j \end{bmatrix}. \tag{6.14}$$

Observe that Γ^{Rj-1} is a vector of belief masses of the leaf elements. The pignistic probabilities of the leaf elements should not be used in Γ^{Rj-1} because of the following reason: At level j, $BetP^j$ is formed using belief masses of composite subsets of the parent frames which are propagated down to subsets of child frames. Therefore $BetP^j$ already takes partially into account the belief masses of the composite subsets of the parent frames. Now computing pignistic probability of a particular leaf element which is a member of the parent frame would again use the same belief masses assigned to the composite subsets (subsets which contain the leaf element) of that parent frame which were used in calculating $BetP^j$. This will lead to a part of the belief mass assigned to a composite subset of a parent frame to be counted twice.

The combined probability vector Pr could also have been obtained from a belief mass assigned to the frame Ψ directly. For any frame such as one defined in (6.4) such a belief mass (here denoted by m^j_Ψ) would have the form shown in (6.15).

The expression in (6.15) is a normalized belief mass assignment similar to any belief mass assigned to the power set of Ψ directly. However, the advantage of using (6.14) is that there is no need to work with the large frame Ψ and its associated

belief mass m^j_Ψ; instead smaller frames and the set V^j at each level can be used to generate partial pignistic probabilities and since the child frames on which the pignistic transformation is applied are disjoint, the partial probability measures can then be finally concatenated together as defined in (6.14).

$$m^j_\Psi(X) = \begin{cases} m_{\Theta^j[P_i]}(X), & \text{if } X \subseteq \Theta^j[P_i], \forall i = 1...k \\ m_Q(Y), & \text{if } X = \cup_i \Theta^j[P_i], i \subseteq \{1, ..., k\} \\ & \text{and } Y = \cup_i P_i, i \subseteq \{1, ..., k\} \\ & \text{and } P_i \in Q \forall i = 1, ..., k \\ m(r), & \text{if } X = r, \forall r \in G^{j-1} \\ 0, & \text{otherwise.} \end{cases} \tag{6.15}$$

The pignistic probability so obtained would be same to what would have been obtained if the transformation in (6.11) was applied on the belief mass m^j_Ψ directly. The need for the vector Pr is to aid in decision making and drawing inferences about the event under observation.

Continuing with the running example, we illustrate the formation of pignistic probability vectors for the consolidated frames mentioned in equations (6.8) and (6.9). At level 1, from each child frame $\Theta^1[H_1], \Theta^1[H_2]$ and $\Theta^1[H_3]$, a pignistic probability vector can be computed for their respective singleton elements. Let these be denoted

as $BetP^{\Theta^1[H_1]}, BetP^{\Theta^1[H_2]}$ and $BetP^{\Theta^1[H_3]}$ respectively where

$$BetP^{\Theta^1[H_1]} = \begin{bmatrix} BetP^{\Theta^1[H_1]}(A) \\ BetP^{\Theta^1[H_1]}(B) \end{bmatrix},$$

with the other vectors formed similarly. As mentioned earlier, the set V^1 is empty since there are no composite subsets of the frame Ω in level 0 with a nonzero belief mass. Also since level 0 has no leaf elements, the set R^0 is empty. Therefore, the net pignistic probability vector for the frames in (6.8) is:

$$Pr = \begin{bmatrix} BetP^{\Theta^1[H_1]} \\ BetP^{\Theta^1[H_2]} \\ BetP^{\Theta^1[H_3]} \end{bmatrix} + \begin{bmatrix} BetP^1 \end{bmatrix},$$

where $BetP^1$ is a zero vector of dimension $|\Psi| \times 1$.

At level 2, the child frames are $\Theta^2[A], \Theta^2[C]$ and $\Theta^2[D]$ and the associated pignistic probability vectors are formed similarly as described before. The composite subsets of parent frames in level 1 whose belief masses would be transferred to elements of V^2 are $\{\{A, B\}, \{C, D\}, \{E, F\}\}$. The masses of these subsets are transferred according to (6.5) and (6.6). The elements of V^2 then would be$\{G, H, B\}, \{I, J, K, L\}$ and $\{E, F\}$ with each having a non zero belief mass. Application of the non normalized pignistic transformation on V^2 would provide probabilistic measures for all the constituent elements of Ψ defined in (6.9). This vector of probability measures forms $BetP^2$.

Therefore, the final net probability vector for the frame in (6.9) is

$$Pr = \begin{bmatrix} BetP^{\Theta^2[A]} \\ BetP^{\Theta^2[C]} \\ BetP^{\Theta^2[D]} \\ \Gamma^{R^1} \end{bmatrix} + \begin{bmatrix} BetP^2 \end{bmatrix},$$

where $\Gamma^{R^1} = \begin{bmatrix} m_{\Theta^1[H_1]}(B) \\ m_{\Theta^1[H_3]}(E) \\ m_{\Theta^1[H_3]}(F) \end{bmatrix}$. Note that, while adding the vectors above, the order must be maintained so that the probabilities of the same elements get added.

6.1.3 Belief Propagation in Evidence Tree: Example

The following example illustrates the application of the ideas developed in the previous sections. Let us assume a hypothetical target classification scenario where a single target is being tracked. Let a CRo be provided against the coarse frame as $\Omega = \{H_0, H_1\}$ where $H_0 = Aircrafts$ and $H_1 = Missiles$ with the refinements defined as $\Theta^1[H_0] = \{Fighter(F), Bomber(B), Cargo(C)\}$ and $\Theta^1[H_1] = \{Ballistic(Ba), Cruise(Cr)\}$. A sample belief mass assignment by a soft source towards Ω can be $m_\Omega(H_0) = 0.6$, $m_\Omega(H_1) = 0.4$ and $m_\Omega(\Omega) = 0$. If in addition to being 60% sure about the target being an aircraft, let the source also claim an opinion of the form *"I am 60% sure about the target being an aircraft and 40% sure about it being a missile but in case its an aircraft, there is high chance it is a fighter"*.

Tables.6.1 and 6.2 show a sample belief mass assignment of the observer toward

Table 6.1: Belief Mass Assignments for $\Theta^1[H_0]$.

Belief Masses	**F**	**B**	**C**	**{F,B }**	**{B,C }**	**{C,F }**	$\Theta^1[H_0]$
Conditional mass ($m'_{\Theta^1[H_0]}(x)$)	0.7	0.2	0	0	0.1	0	0
Scaled mass ($m_{\Theta^1[H_0]}(x)$)	0.42	0.12	0	0	0.06	0	0

Table 6.2: Belief Mass Assignments for $\Theta^1[H_1]$.

Belief Masses	**Ba**	**Cr**	$\Theta^1[H_1]$
Conditional mass ($m'_{\Theta^1[H_1]}(x)$)	0.2	0.7	0.1
Scaled mass ($m_{\Theta^1[H_1]}(x)$)	0.08	0.28	0.04

$\Theta^1[H_0]$ and $\Theta^1[H_1]$ respectively. The pignistic probabilities for the singleton propositions in each refined frame are computed using (6.11) and are shown below

$$BetP^{\Theta^1[H_0]}\begin{bmatrix} F \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0.42 \\ 0.15 \\ 0.03 \end{bmatrix}, BetP^{\Theta^1[H_1]}\begin{bmatrix} Ba \\ Cr \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix}$$

The consolidated pignistic probability vector is then given as

$$Pr = \begin{bmatrix} BetP^{\Theta^1[H_0]} \\ BetP^{\Theta^1[H_1]} \end{bmatrix}$$

Here, $BetP^1$ is a zero vector since the belief mass assigned on Ω is Bayesian. Therefore instead of working with a big frame $\Psi = \{Fighter, Bomber, Cargo, Ballistic, Cruise\}$ and its associated power set which in this example would consist of 32 elements, we can work with much smaller refined frames $\Theta^1[H_0]$ and $\Theta^1[H_1]$. As mentioned before,

(6.15) can provide a normalized belief mass summing up to 1 which would be similar to a belief mass assigned to the power set of Ψ directly.

6.2 Fusion Scenarios

In general, the class of opinions detailed in the previous sections could arise from soft sources especially human observers. This section illustrates situations where multiple such opinions are present with some of them received from hard sensors as well. The end goal is to use informations from all sources to form an updated tree from where pignistic probabilities could be computed to aid in decision making.

6.2.1 Case I: All observers provide opinions toward a fixed tree

This case pertains to the scenario where before the commencement of the fusion process, the general structure of a CRo with various classes and subclasses have been decided and are known such that a tree T has already been formed and various frames placed at its suitable nodes. The observers or sources provide their opinions against this fixed tree T. In other words, the classes and subclasses remain same, however each observer assigns different belief masses to the various frames. The fusion can then be accomplished by using any normalized conjunctive combination operator on the unscaled or conditional belief masses at each level.

For the fusion of a tree across any level, it is preferred if the Dempster's rule of combination is not used. The reason is that Dempster's rule has very high *inertia* when one of the belief masses being combined is categorical with unity mass assigned to some subset which is neither the empty set nor the total frame. By high *inertia*

we mean that combining multiple belief mass assignments with a categorical belief mass assignment using Dempster's rule would result in the combined belief mass to be same as that of the categorical belief mass irrespective of the supports provided by the non-categorical belief masses. All subsequent combinations (no matter how large) would not alter the fused result (assuming there is no belief mass with total conflict). This property in some scenarios may be undesirable. Note that no belief combination operator can be universal in applicability. However, in the context of this discussion, any normalized conjunctive combination operator which does not have an inertia like Dempster's rule operator can be used. For example, the *Weighted Average Operator* (WAO) proposed by Josang et al. in [59] would fit nicely in the belief combination scheme discussed henceforth. This operator performs conjunctive combinations, distributes the conflict proportionally to all the subsets of the frame, and does not provide counter intuitive results under extreme conflict. The fusion steps for a group of N observers are outlined below:

1. Let $m_{\Theta^j}^{i'}$, $i = 1, ..., N$, $j = 1, ..., M$ be the unscaled or conditional belief masses assigned by the i^{th} observer toward a frame Θ^j at level j. For each frame at every level including the root level, a normalized conjunctive combination operator $\odot$ is applied on the unscaled belief masses to obtain

$$m_{\Theta^j}^{(1..N)'} = m_{\Theta^j}^{1'} \odot m_{\Theta^j}^{2'} ... \odot m_{\Theta^j}^{N'}. \tag{6.16}$$

   Note that since all unscaled masses are normalized, the combined result $m_{\Theta^j}^{(1..N)'}$ for every frame will also be normalized.

2. The combined belief masses for all frames in level 1 are scaled using the combined belief masses of the elements of the root frame Ω. In other words, for any frame $\Theta^1[H_i]$ with H_i as the parent in Ω, the scaled mass would be $\forall X \subseteq \Theta^1[H_i]$

$$m^{(1..N)}_{\Theta^1[H_i]}(N) = m^{(1..N)'}_{\Theta^1[H_i]}(X) m^{(1..N)'}_{\Omega}(H_i). \tag{6.17}$$

3. The process of scaling is applied to all subsequent levels until the leaf frames are reached. The scaling of belief masses of elements in a frame at level $j > 1$ are done using the already scaled belief mass of its parent element in level $j-1$. Therefore for a child frame $\Theta^j[P]$ at level $j > 1$ with parent element P in frame Q at level $j-1$, the scaled belief mass would be, $\forall X \subseteq \Theta^j[P]$

$$m^{(1..N)}_{\Theta^j[P]}(X) = m^{(1..N)'}_{\Theta^j[P]}(X) m^{(1..N)}_{Q}(P), \tag{6.18}$$

where $m^{(1..N)}_{Q}(P)$ is the scaled mass of the parent element P in the frame Q. The pignistic probability vector can then be computed using (6.14) from the fused tree.

6.2.2 Case II: Fusion between disparate opinion spaces

In this situation, a tree T is fixed and an observer provides a CRo toward T. Additionally another observer provides an opinion but not against the whole tree but against a single frame Θ. The assumption is that Θ is contained in T at some level with a parent P.

Case IIa: Let us consider the case when the second observer is 100% sure about

the truth being in the class defined by P and the observer distributes the total belief (unity) through a belief mass assigned to the frame Θ consisting of subclasses of P. In such a situation the whole tree can be built with an unity mass assigned to every parent element of the frame Θ all the way up to and including the ancestor in the root frame Ω. All other belief masses will be zero. Once the tree is built using the opinion of observer 2 toward frame Θ, the exact steps outlined in *Case I* could be performed for combining the tree across all levels to compute the fused result.

Case IIb: The assumption here is that the second observer has information only about the frame Θ and is ignorant about any other case in the tree T; by ignorant we mean the observer has no knowledge whatsoever about any of the other cases. Note that this is slightly different than the case of total uncertainty which in evidence theory is represented by assigning a unity mass to the frame. Let the frame Θ be actually at level j of the tree. The following steps could be used to perform the fusion:

1. If the normalized (unscaled) belief masses for the frame Θ provided by observer 2 are given as $m_{\Theta}^{2'}$ then the combination would be performed as

$$m_{\Theta}^{(12)'} = m_{\Theta^{j}[P]}^{1'} \odot m_{\Theta}^{2'}, \tag{6.19}$$

   where $m_{\Theta^{j}[P]}^{1'}$ is the normalized belief mass provided by observer 1 for the frame Θ which is located at level j with a parent P in the tree T.

2. The combined belief masses are scaled $\forall X \subseteq \Theta$, as follows

$$m_{\Theta^{j}[P]}^{(12)}(X) = m_{\Theta}^{(12)'}(X) m_{Q}^{1}(P), \tag{6.20}$$

where $m^1_Q(P)$ is the scaled belief mass of the parent element P. This mass is obtained by representing the CRo of observer 1 using the tree T.

3. Since no information is available from observer 2 about any frame in T other than Θ, the structure and nodes of the tree remain exactly the same as they were before combination. However, the belief masses for all the child frames of the elements of Θ now should be scaled according to the new belief mass $m^{(12)}_{\Theta^j[P]}$. For the child frame of a particular element $\theta_i \in \Theta^j[P]$ this is represented $\forall x \subseteq \Theta^{j+1}[\theta_i]$ as

$$m^{(12)}_{\Theta^{j+1}[\theta_i]}(X) = m^{1'}_{\Theta^{j+1}[\theta_i]}(x) m^{(12)}_{\Theta^j[P]}(\theta_i), \tag{6.21}$$

where $m^{1'}_{\Theta^{j+1}[\theta_i]}$ is the unscaled normalized belief mass provided by observer 1 toward the child frame $\Theta^{j+1}[\theta_i]$. The above operation must be performed for every child frame having parents in $\Theta^j[P]$. Belief masses of elements of all other frames are kept as they were, since no additional information toward those elements was received from observer 2 which would necessitate a change in their belief masses.

In another variation, observer 2 may provide opinions toward a small tree which is a subset of the original tree T. The same process can then be applied as well to perform fusion. The method described in this section can also be treated as a belief revision process since a CRo toward the whole tree must be available before the fusion process begins. Only then opinions toward a single frame or a subtree can

be combined. This makes the process sequential. Once a starting CRo is obtained, subsequent combination can be performed in any order provided the combination operator is associative and commutative.

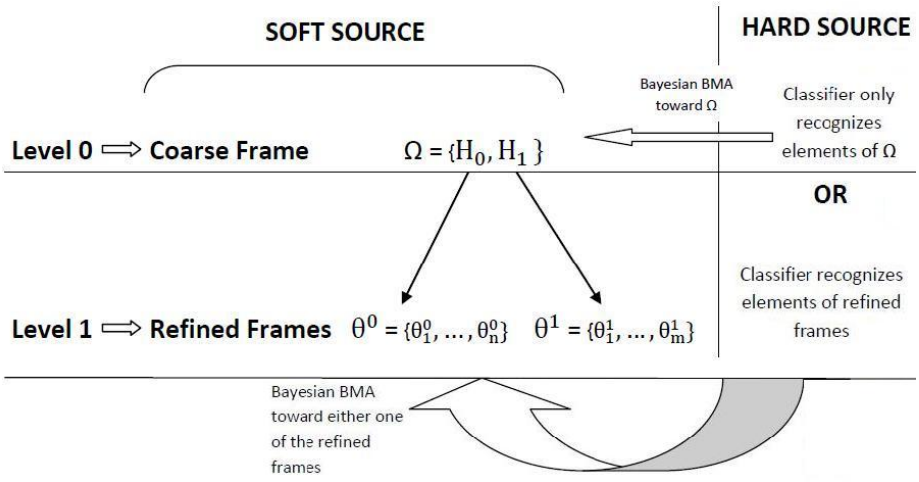


Figure 6.3: Schematic of Hard/Soft fusion scenario with hierarchical evidence trees. (BMA: Belief Mass Assignment)

6.2.3 Case III: Hard/Soft fusion

Let there be a centralized data fusion architecture where a fusion center receives data in the form of belief mass assignments from multiple distributed sensors and implements a fusion algorithm to produce a combined result. For the sake of discussion, let us assume there are two sources which send out belief mass assignments, one is a human observer sending out a CRo and the other source is a hard sensor. Without loss of generality, for the sake of simplicity in explanation, let us assume a 2 level ($M = 1$) tree with a binary root frame given by $\Psi = \{H_0, H_1\}$ and let the corresponding child frames be $\Theta^1[H_0]$ and $\Theta^1[H_1]$. Within this scenario a couple of situations can arise as shown in Fig. 6.3 and explained below:

Case IIIa: The hard sensor classifier can only distinguish between elements of

the coarse frame Ω: In this case, the classifier output will be a Bayesian belief mass assignment towards Ω. We say Bayesian belief mass assignment, since we assume the classifier sends out probabilities of classification and that pertains to only the atomic subsets of a frame.

Case IIIb: The hard sensor classifier can distinguish between the elements of the refined frames: The classifier output will be a Bayesian belief mass assignment toward either of the refined frames. Here we assume, if a classifier has narrowed down to one of the refined frames, then it is 100% certain about the class in the coarse frame. In other words, if a classifier sends out a belief mass assignment toward $\Theta^1[H_0]$, then it means that the elements of $\Theta^1[H_1]$ which in turn belong to the class defined by H_1 are ruled out. Therefore, the classifier output can be a Bayesian belief mass assignment toward either of the refined frames and not both. The two cases are elucidated in detail as follows:

Case IIIa: Let the belief mass assignments toward the coarse frame Ω be denoted by m_{Ω_s} and m_{Ω_h} for the soft and hard sensors respectively where m_{Ω_h} is assumed to be Bayesian. The human observer also provides normalized belief mass assignments $m'_{\Theta^1[H_0]}$, $m'_{\Theta^1[H_0]}$ toward the refined frames $\Theta^1[H_0]$ and $\Theta^1[H_1]$. We outline the fusion steps below:

1. Combine m_{Ω_s} and m_{Ω_h} using any normalized conjunctive combination rule to form $m_{\Omega_{sh}} = m_{\Omega_s} \odot m_{\Omega_h}$ where $\odot$ is the combination operator.

2. Scale the normalized belief mass assignments $m'_{\Theta^1[H_0]}$ and $m'_{\Theta^1[H_0]}$ as follows

$$m_{\Theta^1[H_i]}(X) = m'_{\Theta^1[H_i]} m_{\Omega_{sh}}(H_i),$$

$\forall X \subseteq \Theta^1[H_i], X \neq \phi$ where $i = 0, 1$. The scaled belief masses would then satisfy

$$\sum_{X \subseteq \Theta^1[H_0]} m_{\Theta^1[H_0]}(X) = m_{\Omega_{sh}}(H_0)$$
$$\sum_{X \subseteq \Theta^1[H_1]} m_{\Theta^1[H_1]}(X) = m_{\Omega_{sh}}(H_1).$$

3. Once the updated tree is available, either the combined belief mass over the consolidated frame Ψ can be obtained using (6.15) or a pignistic probability vector can be computed for subsequent decision making. The pignistic probability vector can be formed by using the transformation defined in (6.11) on the masses $m_{\Theta^1[H_0]}$ and $m_{\Theta^1[H_1]}$ to form $BetP^{\Theta^1[H_0]}$ and $BetP^{\Theta^1[H_1]}$ respectively. The consolidated pignistic probability vector would then be given by (6.14). If m_{Ω_s} is not Bayesian , then $BetP^1$ will have nonzero values or else it would be a $|\Psi| \times 1$ zero vector where $|\Psi| = |\Theta^1[H_0]| + |\Theta^1[H_1]|$.

Case IIIb: In this situation, the hard sensor classifier can distinguish between elements of the refined frames and hence sends out a Bayesian belief mass assignment toward any one of the refined frames. As the hard sensor sends out the Bayesian belief mass toward one of the refined frames, it implicitly means the classifier is 100% sure about which class of the coarse frame is true. In other words, if the classifier sends out a Bayesian belief mass set toward $\Theta^1[H_i]$ which is a refinement of H_i, then

Table 6.3: Combined support values for hierarchical hard and soft opinions.

Propositions	**Combined belief mass**	**Pignistic Probability**
F	0.469	0.469
B	0.386	0.400
C	0.037	0.051
{F,B }	0	-
{B,C }	0.028	-
{C,F }	0	-
$\mathbf{\Theta^1[H_0]}$	0	-
Ba	0.016	0.020
Cr	0.056	0.060
$\mathbf{\Theta^1[H_1]}$	0.008	-

the following is true

$$\begin{aligned} m_{\Omega_h}(H_i) &= 1 \\ m_{\Omega_h}(X) &= 0, \forall X \subseteq \Omega, X \neq H_i. \end{aligned}$$

Let the Bayesian belief mass assignment obtained from the hard sensor classifier be toward the refined frame $\Theta^1[H_i]$ and let it be defined as $m_{\Theta^1[H_i]_h}$. The situation then becomes identical to *Case IIa* described above. The fusion process therefore could be performed similarly. All the above methods of fusion would suit well for hard soft data fusion and are compatible with algorithms proposed in [77].

Example: We continue with the hypothetical example discussed in Section 6.1 and modify it to illustrate the Hard-Soft fusion steps. The soft source belief mass assignments both normalized and scaled are as provided in Tables 6.1 and 6.2. Additionally, let there be a radar (hard sensor) whose classifier decides the target to

be an aircraft and assigns a Bayesian belief mass to the refined frame $\Theta^1[H_0] = \{Fighter(F), Bomber(B), Cargo(C)\}$ given as $m_{\Theta^1[H_0]_h}(F) = 0.3$, $m_{\Theta^1[H_0]_h}(B) = 0.6$ and $m_{\Theta^1[H_0]_h}(C) = 0.1$.

A Bayesian belief mass assignment toward $\Theta^1[H_0]$ means the radar rules out the possibility of H_1 and therefore all elements of its refinement $\Theta^1[H_1]$. In that case, the tree when filled up with the data from the radar would have the belief mass provided above assigned to the child frame $\Theta^1[H_0]$ and the categorical belief mass $m_{\Omega_h}(H_0) = 1$ and $m_{\Omega_h}(H_1) = 0$ assigned to the root frame Ω. The frame $\Theta^1[H_1]$ is assigned zero belief mass. The assignment of zero belief mass is considered as no information. Now following the steps outlined for *Case IIa* the belief masses from both hard and soft sources are combined using the WAO across each level. The resultant combined belief mass for $\Theta^1[H_1]$ is then scaled accordingly. The combination process in level 0 assigns a nonzero belief mass to the class H_1 which was not supported by the hard sensor. Therefore, this resultant nonzero mass is used to scale the normalized belief mass toward $\Theta^1[H_1]$ that was provided by the soft source since no new information about $\Theta^1[H_1]$ was received from the radar and hence the old information must be retained. The final pignistic probabilities are calculated using the scaled belief masses. The final results are provided in Table 6.3. Note that in this example, $BetP^1$ is a zero vector. The pignistic probabilities could now be used for decision making, like in the context of this example, it can be inferred that the target is most likely a fighter.

6.3 Complexity Analysis

Let there be s elements in the root frame Ω of a M level tree (including level 0). Assuming every element in the tree starting from the elements of the root frame to the elements of the frames in level $M-1$ has k children; starting from level 0 downwards, the consolidated frame Ψ at every level has $s, sk, sk^2...$ elements. Hence with increase in levels of the tree, the size of the consolidated frame increases geometrically with k as the scale factor. However, in the proposed scheme, both belief mass and pignistic probability calculations for any consolidated frame Ψ at any level are performed using the individual child frames in that level with the final result being given by a simple concatenation operation. Therefore, computational complexity of the proposed algorithms increases with the tree size only linearly (the operations performed for a small frame needs to be repeated as many times as the number of frames in a particular level) but the real intensive calculations which involve summations or products over all subsets of a frame (pignistic probability and conjunctive belief combination) are performed over small child frames with small number of subsets. The essence of the scheme is thus to substitute multiple implementations of the intensive computations over small frames for a single implementation over a much larger consolidated frame. Hence, if the tree is formed using small sized frames, then the depth of such a tree would not affect the implementation of the proposed scheme.

7: HARD/SOFT FUSION: DISCUSSION AND FUTURE WORK

The belief theory framework was originally introduced by Shafer in [16] as a method of modeling uncertain evidence. The assignment of belief masses to sets of propositions and building a support function over a power set handle ignorance and uncertainty of sources in an intuitive way. Such a modeling method seems very apt to be used to model opinions coming in from human observers as such data lacks the well defined structure required for the use of traditional probabilistic reasoning.

In [77], we proposed an approach based on evidence combination that is able to take in data in the form of probabilistic support values toward a set of hypotheses from both hard and soft sensors and combine such data for informed decision making. Based on an event and available knowledge of its possible outcomes, soft sensor data is modeled using evidence theory. In the case of more than one soft sensor, the consensus operator [60] is used to combine multiple belief functions. After processing hard sensor data to obtain probabilistic support values for various propositions, the fusion of probability rule [26] is implemented to perform final combination of hard and soft data.

Depending upon the situation and the nature of hard sensors employed, the technique to process sensor data may change but the fusion rule would remain the same. Since there are roles for domain experts and analysts like experience and intuition based decision making which are extremely difficult to imitate in a computer algo-

rithm, the whole fusion process does require human intervention and therefore is a semi-automatic procedure. Although the algorithm is easy to implement, the employment of large numbers of humans for testing can be logistically challenging and expensive, and opportunities to re-test the same humans on modified data presentations and exposition schemes is often difficult. At least in the early stages of testing and tuning of data fusion algorithms, it may be desirable to use models of human decision making rather than using actual human-generated data. A few studies in this direction, which use cognitive psychology models like the two stage dynamic signal detection model [89] to simulate human decision making have been reported in [90, 91].

The study in [92] developed a hierarchical evidence tree structure to represent a special class of nested human opinions call Conditionally Refined opinions (CRo). Such structures were shown to be able to model equivocal and nested human opinions. The representation was also shown to be compatible with any belief theoretic soft-soft and hard-soft fusion methodologies including the scheme proposed in [77]. However the model in [92] is static in the sense that data pertaining to new classes cannot be accommodated in an existing tree. Future research can be directed to extend this work and develop a dynamic tree structure which can aid in sequential fusion of human opinions. In other words, the tree would evolve temporally with addition of new nodes and removal of old and irrelevant nodes. A more complex idea of rearranging new and old nodes based on variable hierarchical relationships could also be investigated.

Further effort can be directed to develop a generic Hard/Soft fusion system that

would be almost automated and be able to take in subjective informations from humans and also data from hard sensors and integrate them in a computationally efficient manner. Soft (human opinion) data can be collected from surveys or polls created through crowd sourcing (e.g., Amazon Mechanical Turk) and parsed using customized text parsers for further processing and fusion with hard data.

Part III

APPLICATION OF PARALLEL DISTRIBUTED DETECTION AND FUSION

8: ACTIVE AUTHENTICATION WITH BIOMETRIC SENSORS

The interaction between humans and most desktop and laptop computers is often performed through two input devices: the keyboard and the mouse. Continuous tracking of these devices provides an opportunity to verify the identity of an user, based on a profile of behavioral biometrics from the user's previous interaction with these devices. We consider the real-time application of this technology for active authentication. As a user begins interacting with the machine, the classification system collects behavioral biometrics from the interaction and continuously verifies that the current user has access permission on the machine. This approach adds an extra layer of distraction-less access control in environments where a computer is at a risk of being intermittently accessed by unauthorized users.

We propose a bank of sensors, each feeding a binary detector (trying to distinguish the authentic user from all others). In this study the detectors use features derived from the keyboard and the mouse, and their decisions are fused to develop a global authentication decision. The binary classification of the individual features is developed using Naive Bayes Classifiers which play the role of local detectors in a parallel binary decision fusion architecture. The conclusion of each classifier ('authentic user' or 'other') is sent to a decision fusion center where we use the Neyman-Pearson criterion to maximize the probability of detection under an upper bound on the probability of false alarms. We compute the receiver operating characteristic (ROC) curves of

the resulting detection scheme, and use the ROC curves to assess the contribution of each individual sensor to the quality of the global decision on user authenticity. In this manner we identify the characteristics (and local detectors) that are most significant to the development of correct user authentication. While the false alarm and mis-detection rates are fixed for the local sensors, the fusion center provides trade-off between the two global error rates, and allows the designer to fix an operating point based on his/her tolerance level of false alarms. We test our approach on a real-world dataset collected from 10 office workers, who worked for a week in an office environment as we tracked their keyboard dynamics and mouse movements during interaction with laptops and desktop computers.

8.1 Context of Active Authentication

The tracking of behavioral biometrics for continuous verification of an user's identity has received considerable attention in recent years [93]. By monitoring actively metrics such as keyboard dynamics and mouse movements, classification of user as authentic or non-authentic has achieved accuracy on par with more traditional non-continuous approaches [94]. One popular non-continuous approach is for the user to verify his/her identity by typing a password or a common fixed phrase. The authentication system then estimates whether the user is who s/he claims to be by analyzing the biometric parameters associated with the typing of the password/phrase. Continuous "active authentication", on the other hand performs verification of the user steadily, based on a set of metrics collected during previous interaction with the computer, or updated based on known-user behavior. Due to the unconstrained nature

of human-computer interaction, a single biometric is usually not sufficiently robust to determine the user's identity. For that reason, many active authentication systems are (a) multi-modality, namely they monitor multiple features of a single type of biometric [95], and (b) multi-biometric, namely they consider more than one type of biometric [96]. In this study we consider multi-modality multi-biometric model for interaction with a computer through a mouse and keyboard.

We evaluated our algorithms using a dataset collected from office workers in a real-world office environment. Each user is represented through features collected from the user's keyboard dynamics and mouse movements. We fuse these features using established algorithms for parallel binary decision fusion [20, 97]. The Receiver Operating Characteristic (ROC) helps quantify the relative importance of each biometric and each feature.

8.2 User Authentication via Biometrics

8.2.1 Mouse and Keyboard Dynamics

The movement dynamics of the mouse and the keyboard of a personal computer have been studied for over two decades [98, 99] as the primary human computer interface input devices. Keystroke dynamics have received most of the attention in behavioral biometrics studies [100]. The two basic tracked features of inter-key-press interval [101] and key-press dwell [102] were used as the basis for more complex features such as digraph latency [103], trigraph latency [104], or keyword latency [105]. These features provide timing information about a specific configuration of key-press and key-release events.

Mouse movement in the authentication domain has received considerably less attention until recently [106]. Mouse movement signals are relatively noisy, requiring large testing windows for authentication [107]. Zheng et al. [108] were the first to show meaningful results based on testing windows as small as 20 mouse clicks. Jorgensen and Yu [109] studied short-time window continuous authentication and used touch-based mouse devices.

Most single-modality classifiers considered for mouse and keyboard use statistical methods [110] such as Naive Bayes, decision trees, linear discrimination analysis and support vector machines [111]. Some classifiers use trained neural networks [112]. In this work, our local detectors are Naive Bayes classifiers [113] due to their robustness to the relatively small amount of training data needed and the ease of design and implementation.

8.2.2 Multi-Biometric Systems

In the context of continuous authentication, multi-modality biometrics use multiple asynchronous streams of features to form a sequence of fused verification decisions. There are several approaches to categorize multi-biometric systems [114]. Most relevant to the approach in this study is division of continuous fusion by Sanderson and Paliwal into: (1) pre-mapping, (2) midst-mapping, and (3) post-mapping [115]. Here "mapping" refers to the transformation from data into information (e.g, from the feature space to the decision space). While pre-mapping fusion has been extensively studied in other applications [72], in the authentication domain, the feature space is so varied (and often restricted by privacy concerns) that midst-mapping or post-

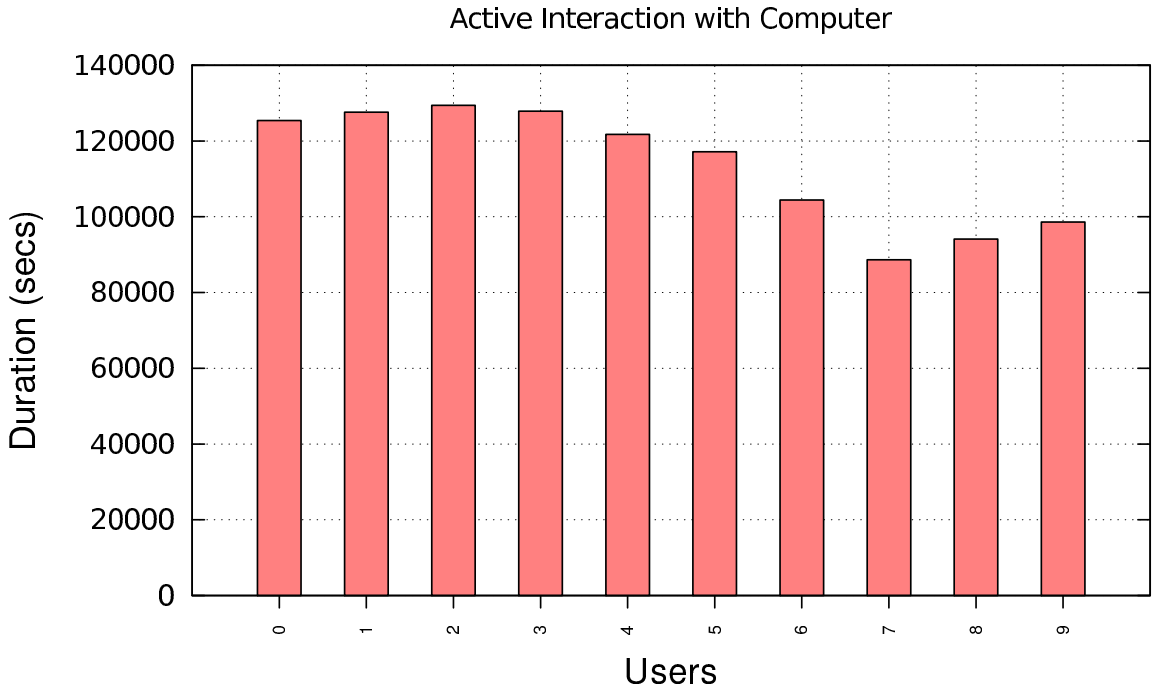


Figure 8.1: The duration (in seconds) of each user's interaction with the computer throughout the 5-day week with idle periods removed. An idle period is defined as a continuous period of time without any mouse or keyboard interaction with the computer.

mapping fusion have been preferred. We use the post-mapping (or decision fusion) approach in this study.

8.3 Experimental Setup

8.3.1 Dataset

The dataset used in our study comes from 10 users in a simulated work environment. The users were tracked throughout a working week (5 sequential days) in their use of the mouse and keyboard as they sought to accomplish various writing tasks such as summarizing on-line opinion articles. The productivity, task-selection, and mouse/keyboard use ratio varied from user to user as shown in Figure 8.1 and

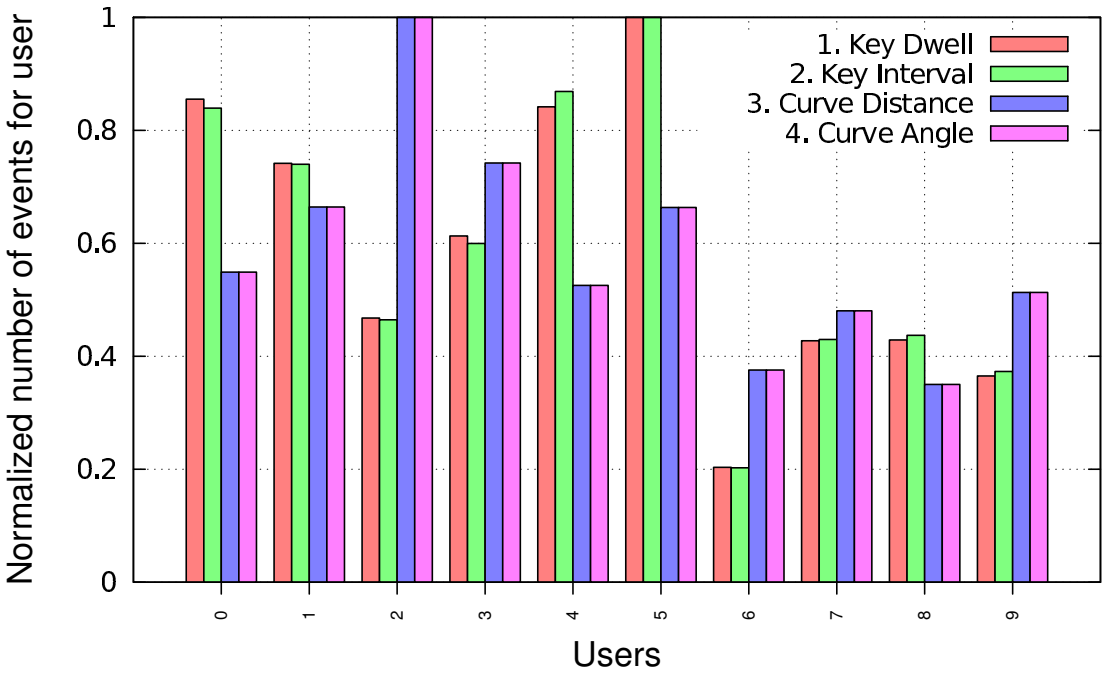


Figure 8.2: The relative amount of biometric data per-type per-user extracted from the interaction of each user with their computer throughout the 5-day week. The variability between the users is noticeable.

Figure 8.2.

A tracking application logged two types of behavioral biometrics on the granularity of 5 milliseconds:

- Mouse movement, mouse click, and mouse scroll wheel events.
- Keystroke dynamics (include press, hold, release durations) for all keyboard keys including special keys.

Figures 8.3 and 8.4 illustrate the mouse and keystroke dynamics features.

Table 8.1 shows statistics on the biometric data in the dataset. The table contains data aggregated over all 10 users and all of the available user-days. The

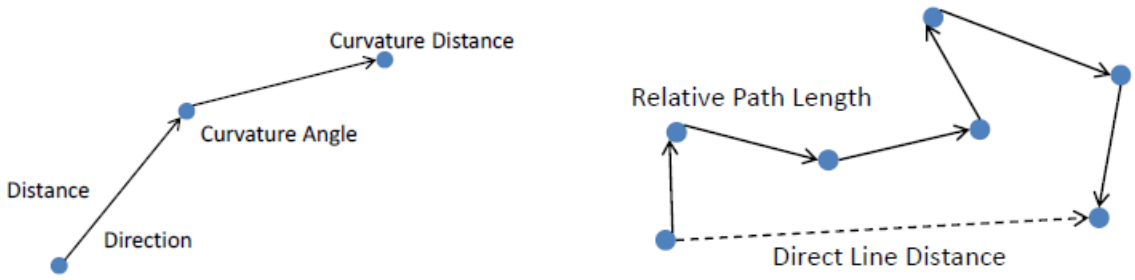


Figure 8.3: The mouse movement metrics are computed from a set of continuous move events.

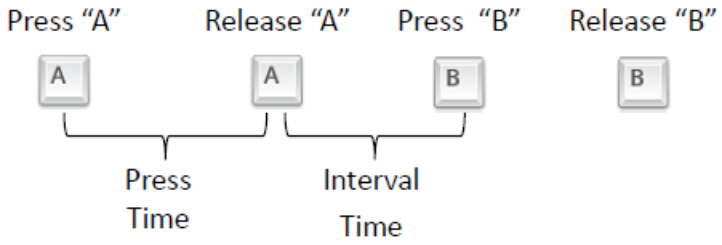


Figure 8.4: The keystroke dynamics metrics are computed from time between the press and the release event and vice versa.

keystroke events include both the alpha-numeric keys and special keys such as `shift`, `backspace`, `ctrl` and `alt`.

Table 8.1: Statistics on the 10-user dataset.

Metric	**Total**
Sensor 1: Keystroke Dwell Time	915,624
Sensor 2. Keystroke Interval Time	750,253
Sensor 3: Mouse Curvature Distance	3,462,912
Sensor 4: Mouse Curvature Angle	3,462,912

Table 8.1 shows the total quantity of features extracted from the raw data for each of the four sensors. The number of keystroke intervals is significantly less than keystroke dwell time because only intervals that were part of bursts of continuous typing were collected.

8.3.2 Feature Classification

Data from each user were broken down into three segments relative to the duration of each user's non-idle periods of activity (see Figure 8.1):

1. *Training Segment:* for the construction of the empirical feature distribution;

2. *Characterization Segment:* for the estimation of the false alarm and mis-detection rates of the sensor for the use by the decision fusion center; and

3. *Testing Segment:* for the estimation of the local sensor performance compared to the fused sensor performance.

For each user we trained the classifier on the first 60% of the data, characterized the error rates of the trained classifier on the following 20%, and tested the local and the fused sensors on the remaining 20%.

For each local sensor, we use the Naive Bayes Classifier [116] for mapping from the feature space to the decision space. This classifier is constructed during the training phase and used for binary classification in the characterization and testing phases.

In the training phase, the empirical distribution for feature probabilities are constructed from the frequency of each feature in the training segment of each user's data. An example histogram-based distribution for the "keystroke dwell time" metric for the first user is shown in Figure 8.5. It shows the estimated probability of dwell time (in seconds). Two such histograms are constructed for each user j. The first histogram was constructed from the training segment of the data of that user. The second histogram was constructed from all the training segments of the other

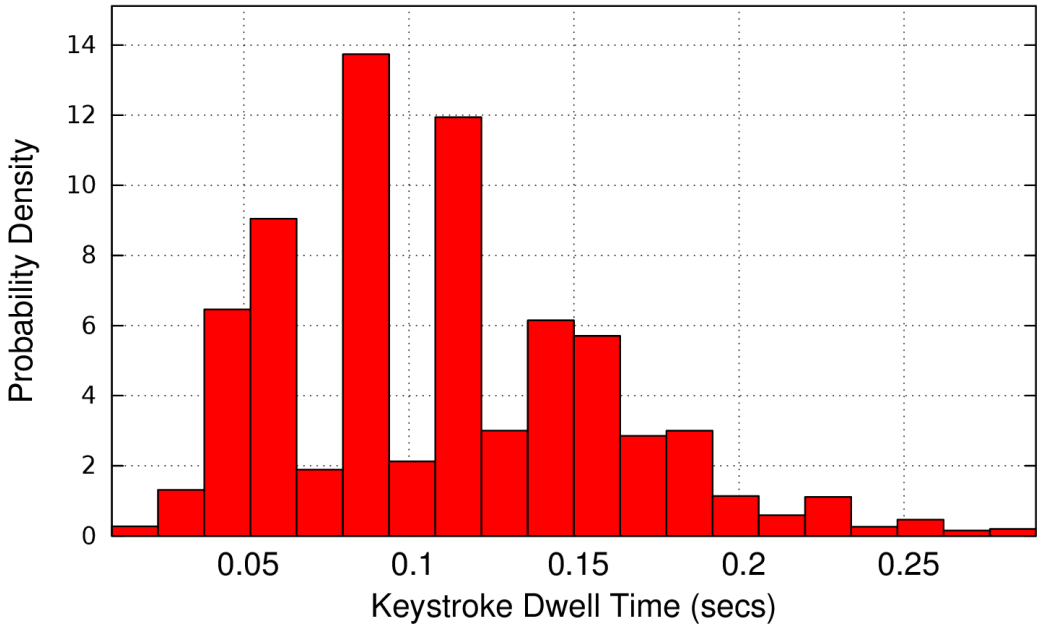


Figure 8.5: An example of a histogram constructed from the training set for the empirical probability distribution of user 0 for the "Keystroke Dwell Time" feature (sensor 1).

users. This latter set of training segments is meant to represent all other computer users. These two histograms are the empirical feature distributions associated with each user.

In the characterization and testing phases, for each user and each metric, the Naive Bayes Classifier considers a collection of ζ (set to 10 in this study) most recent events $\{z_1, z_2, ..., z_\zeta\}$ associated with that metric (e.g, keystroke dwell time). It then uses the maximum a posteriori (MAP) rule to pick the most likely hypothesis:

$$H^* = \underset{i \in \{0,1\}}{\operatorname{argmax}} P(H_i) \prod_{k}^{\zeta} P(z_k | H_i), \tag{8.1}$$

where H_1 is the "authentic" class, H_0 is the "non-authentic" class, discussed further in §8.4, and H^* is the most likely class associated with the observed biometric data.

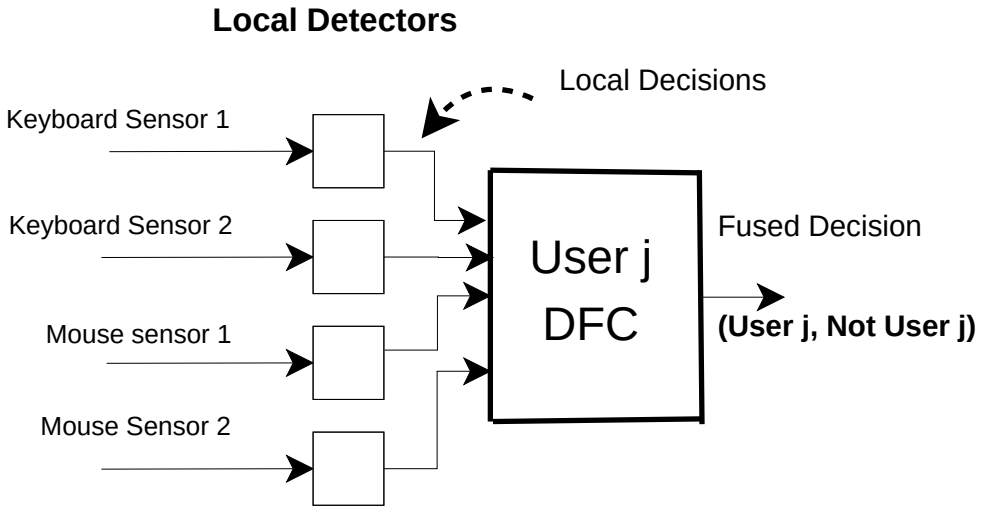


Figure 8.6: Behavioral sensor fusion scheme for active authentication.

Given the open-world model considered in this paper, for the local sensor classifiers, we considered $P(H_0) = P(H_1) = 0.5$. The feature probability $P(z_k|H_i)$ is estimated by a non-parametric distribution (like the one in Figure 8.5).

8.4 Decision Fusion

We employ a decentralized parallel binary decision fusion [20, 97] scheme to integrate the set of local binary decisions to a global binary decision. The local detectors are designed as binary Naive Bayes Classifiers each attached to a single feature of the mouse or the keyboard. The i^{th} local detector is characterized by its probability of correct detection ($P_{D_i} = 1 - P_{M_i}$; and its probability of false alarm (P_{F_i}). Here we selected to design a decision fusion center that integrates the local decisions in the Neyman-Pearson sense [97]: for an upper bound on the global probability of false alarm, the fusion center maximizes the global probability of detection. The distributed decision fusion scheme is depicted in Figure 8.6. It involves 4 different

sensors (2 Mouse and 2 Keyboard) shown in Table 8.1, each connected to a local detector that announces whether or not the user is authentic. The experimental setup involved collecting data from 10 users and creating 10 schemes (local detectors and a Decision Fusion Center), each one designed to identify one of the 10 users and reject all others. The j^{th} fusion center, $(j = 1, ..., 10)$ distinguishes the hypothesis (H_1 = User j is present; against H_0 = User j is not present).

Each sensor, designed as a Naive Bayes Classifier uses its own observations to decide on the hypothesis for the corresponding user. The local decisions are of the form

$$u_i = \begin{cases} 1, & \text{if } H_1 \text{ is accepted (user is authentic)} \\ -1, & \text{if } H_0 \text{ is accepted (user is non-authentic)} \end{cases} \tag{8.2}$$

The fusion center performs a Neyman-Pearson test [117] for fusion and for each user it functions the same way; the fusion center for the j^{th} user takes in the four local decisions and calculates the likelihood ratio for N sensors (in this study, $N = 4$);

$$\Lambda(u^j) \;=\; \frac{P(u_1^j, ..., u_N^j|H_1)}{P(u_1^j, ..., u_N^j|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_g^j. \tag{8.3}$$

Assuming that the local decisions are independent (conditioned on the hypothesis), we have

$$\Lambda(u^j) \;=\; \prod_{j=1}^{N} \frac{P(u_1^j|H_1)}{P(u_1^j|H_0)} \times ... \times \frac{P(u_n^j|H_1)}{P(u_N^j|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_g^j, \tag{8.4}$$

where the threshold t_g^j is computed such that the global false alarm at the fusion

center for the j^{th} user ($P_{F_0}^j$) is α_j. In other words, t_g^j is obtained such that

$$\sum_{\Lambda(u^j)\geq t_g^j} P(\Lambda(u^j)|H_0) = \alpha_j. \tag{8.5}$$

Once the threshold t_j^* is determined, the global probability of detection ($P_{D_0}^j$) at the fusion center for the j^{th} user becomes

$$P_{D_0}^j = \sum_{\Lambda(u^j)\geq t_g^j} P(\Lambda(u^j)|H_1). \tag{8.6}$$

As the global conditional distributions are discrete, only certain global false alarm values would be possible which would make the receiver operating characteristics contain a collection of disjoint points. To allow realization of an intermediate false alarm rate, the fusion center employs a randomized Neyman-Pearson test and computes the probability of using (one of the two) thresholds accordingly.

As the specified global false alarm rate P_{F_0} varies from 0 to 1, the global detection rate ($P_{D_0} = 1 - P_{M_0}$) is calculated to create the Receiver Operating Characteristic (ROC) curves.

8.5 Behavioral Sensor Fusion Performance

The fusion setup used for generating the ROC curves is shown in Figure 8.6. The 4 sensors are assumed to be fixed Naive Bayes Classifiers and the fusion center performs a Neyman-Pearson test to fuse the local detector decisions. A total of 10 users and 10 systems were tested. To compare the contributions of each sensor towards the fused

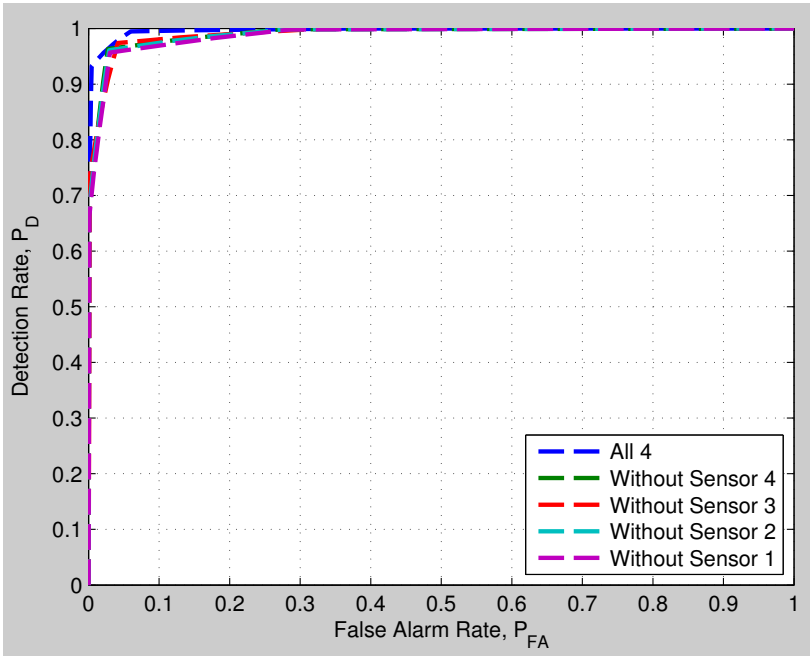


Figure 8.7: ROC curves for incremental and global sensor fusion with one biometric sensor taken out at a time.

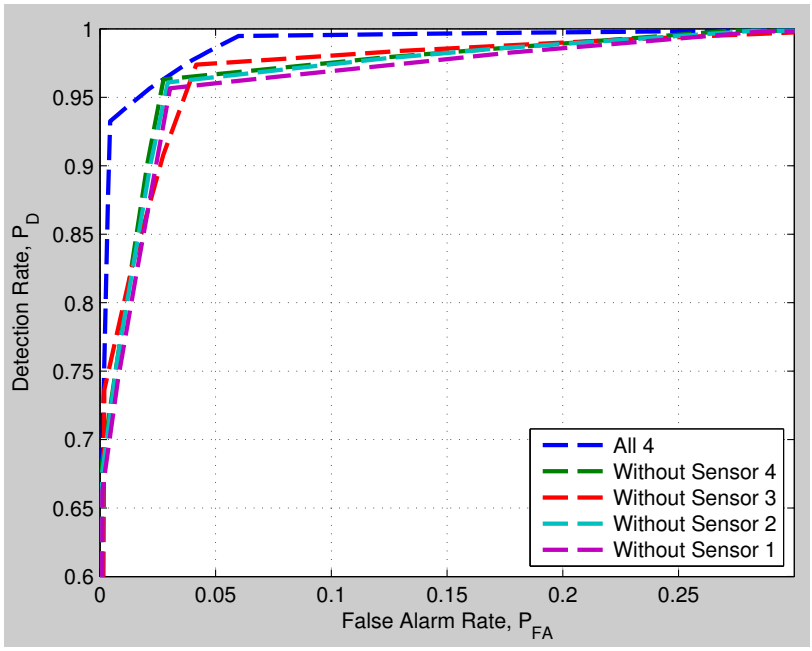


Figure 8.8: Zoomed in version of Figure. 8.7.

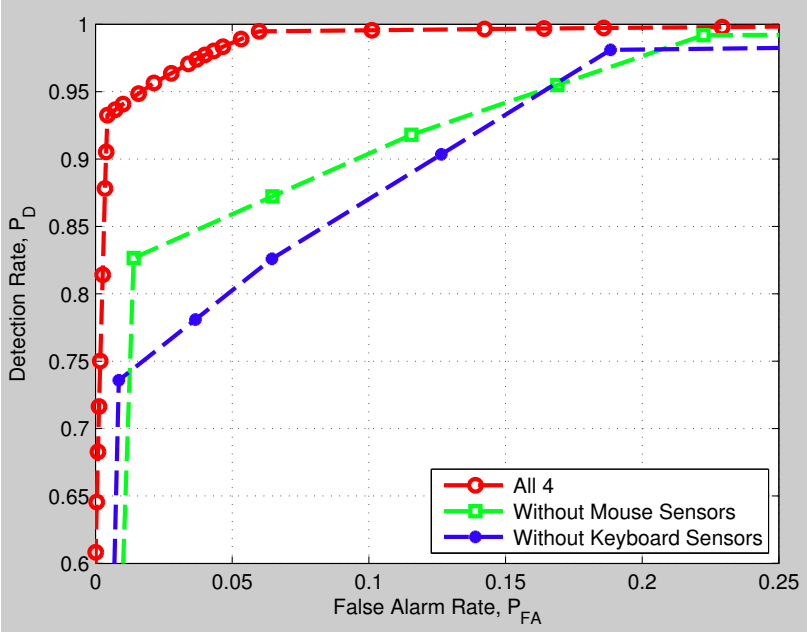


Figure 8.9: ROC curves for incremental sensor fusion with mouse and keyboard sensors removed at a time.

result, each sensor was removed (one at a time) and the ROC curve of the resultant system generated.

Figure 8.7 shows the ROC for user 1. Figure 8.8 shows a zoomed-in version of same plot. It is noticeable that fusion is profitable and produces the highest possible global detection rates for given false alarm rates. All the sensors are very similar in performance and therefore have comparable marginal contribution to the global performance. Similar results were obtained for the decision fusion centers designed for the remaining 9 users. In general this approach of using ROC curves for sensor analysis can help in creating a hierarchy of sensor importance such that for a known tolerable global false alarm, the appropriate groups of sensors could be identified and used (and some sensors of only marginal contribution can be dropped).

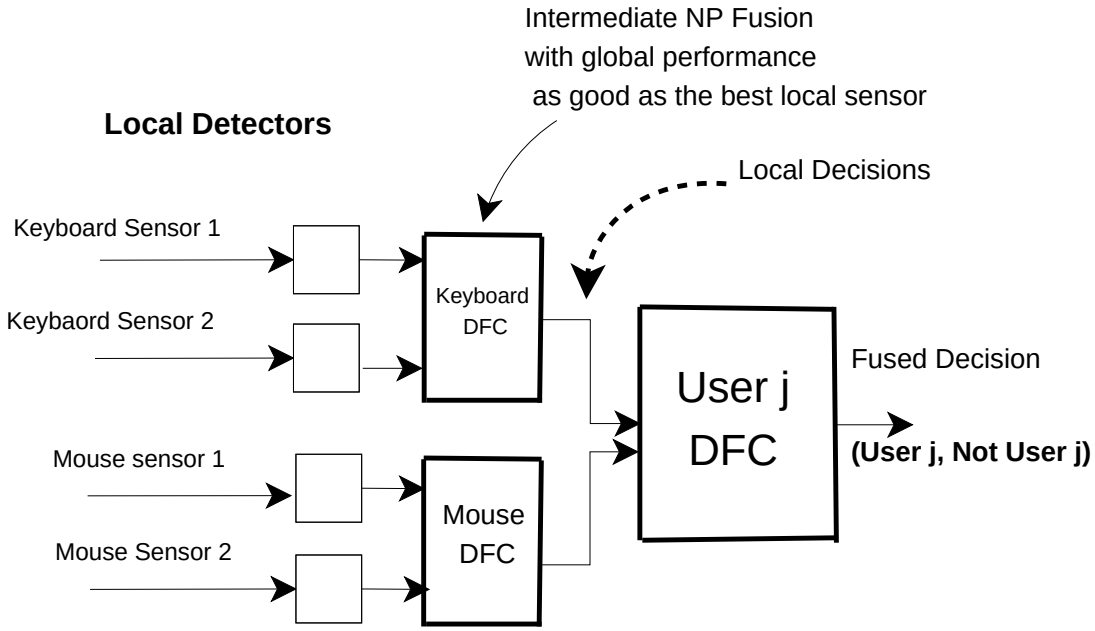


Figure 8.10: Multilevel decision fusion.

Figure 8.9 shows the resultant ROC curves when either the two mouse sensors or the two keyboard sensors were removed incrementally. For small values of global false alarm rates, it is observable that removing the keyboard sensors and combining just the mouse sensors degrades the global performance more than removing the mouse sensors and combining only the keyboard sensors.

8.5.1 Multilevel Fusion

We may ask, how much would performance be affected when decisions associated with the same kind of sensors are combined first and then the results of such intermediate fusions are fused by the main fusion center. In the fusion scheme shown in Figure 8.10 all the mouse and keyboard sensors, were combined first, separately, using Neyman-Pearson criterion; the outputs were then fused at the main fusion center. The false alarm at the intermediate fusion centers were chosen the same as the false alarm of

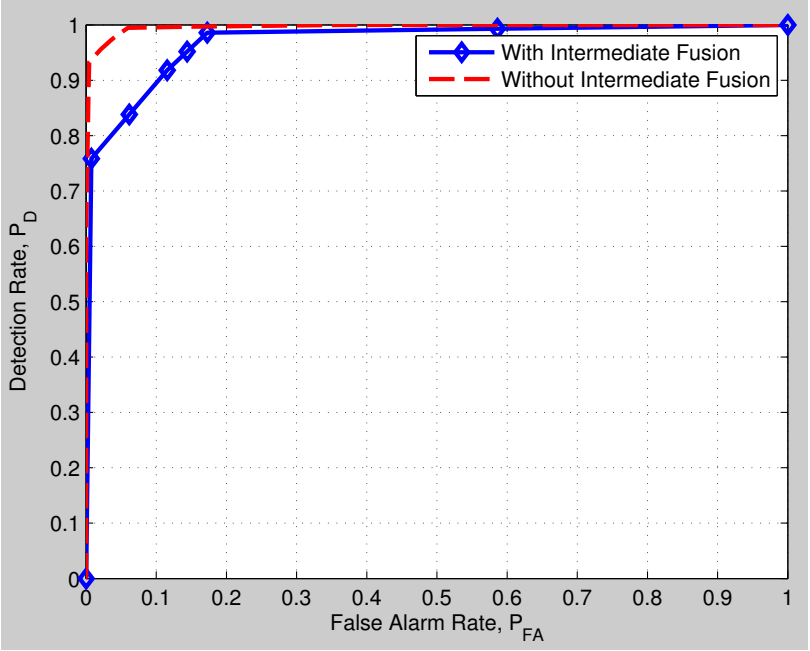


Figure 8.11: Receiver Operating Characteristic for two step fusion.

best individual sensor of that class. For user 1, Figure 8.11 shows the resultant ROC curves at the user fusion center.

Not surprisingly, performing intermediate fusion degrades the overall system performance for low values of false alarm rate.

8.6 Additional Modalities for Active Authentication

In few separate studies [118–120] we extend the work in [121], and consider real time user authentication using *high level* biometrics such as *stylometry* and *web browsing*, in addition to low level mouse and keyboard features. Specifically we employ four classes of biometrics: keystroke dynamics, mouse movement, web browsing and stylometry. Stylometric analysis, in particular is well developed. However, its application to continuous verification of user identity is new. The basic assumption behind

stylometry is that every person has a unique linguistic style (*stylome*, [122]) that can be quantified and measured in order to disntiguish between different authors.

Depending on the what task the user is engaged in, some of the biometric sensors may provide more data than others. For example, as the user browses the web, the mouse and web browsing sensors will be actively flooded with data, while the keystroke dynamics and stylometry sensors may only get a few infrequent updates. This observation motivates the application of fusion under a distributed topology to gain continuous monitoring of users.

8.6.1 Suite of High and Low Level Biometrics

We collected behavioral biometrics data from the same simulated work environment described in Section 8.3.1. However, due to the incorporation of linguistic modalities like stylometry, a larger data set was collected from 5 users each working for 40 hours per week over a duration of 4 weeks. This larger data set was essential for training of support vector machines used to perform classification based on linguistics. The following suite of biometric sensors was used during data collection.

- Low-level sensors:
  - M1: mouse curvature angle
  - M2: mouse curvature distance
  - M3: mouse direction
  - K1: keystroke interval time
  - K2: keystroke dwell time

- High-level sensors:
  - W1: website domain visit frequency
  - S1: stylometry (1000 char., 30 min. window)
  - S2: stylometry (500 char., 30 min. window)
  - S3: stylometry (400 char., 10 min. window)
  - S4: stylometry (100 char., 10 min. window)

8.6.2 Feature Sets and Classification

For stylometry, the feature set used was a variation of the *Writeprints* [123], which includes a vast range of linguistic features across different levels of text. This rich linguistic feature set is aimed at capturing the users writing style. With the special character placeholders, some features capture aspects of the users style usually not found in standard authorship problem settings. For classification we used sequential minimal optimization (SMO) support vector machines with polynomial kernel, available in Weka [124]. Support vector machines are commonly used for authorship attribution [125] and documented to achieve high performance and accuracy.

The low-level metrics of keystroke and mouse dynamics detectors, along with the domain visit frequency detector, all use support vector machines.

We used the same SVM classifier as for low-level sensors, and the feature vector of the visit frequency to the 20 most visited websites in the dataset, the top five of which were: google.com (7.0%), bing.com (7.0%), facebook.com (5.0%), yahoo.com (4.1%), and wikipedia.org (2.9%). The visit frequency of any one of these popular websites is

W1	S1	S2	S3	S4	M1	M2	M3	K1	K2	FAR	FRR
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.00122	0.00218
	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0021	0.0027
✓					✓	✓	✓	✓	✓	0.00728	0.00714
					✓	✓	✓	✓	✓	0.0102	0.01016

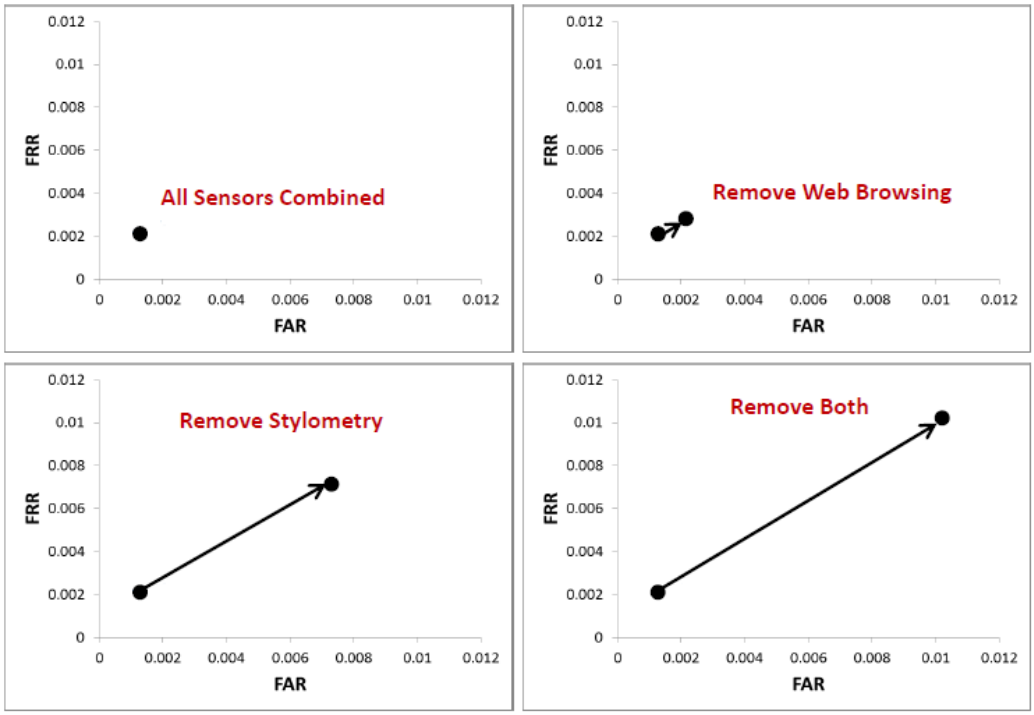


Figure 8.12: False alarm rates (FAR) and mis-detection rates (FRR) for 4 representative selection of sensors of the 1024 possible combinations for fusion. These four cases are: (1) all high and low level modalities are used; (2) all modalities except for web browsing are used; (3) all modalities except for stylometry sensors are used; (4) all modalities except for web browsing and stylometry are used.

not a good classification feature. However, taken together, the 20 dimensional feature vector forms a sufficiently representative profile of a user to be used in continuous authentication.

8.6.3 Fusion of High and Low Level Biometric Features

The same parallel binary decision fusion architecture as described in Section 8.4 was applied, however with the representative collection of both high and low level

behavioral biometric sensors.

The rows of the table in Figure 8.12 are four representative combinations of the 10 sensors listed in Section 8.6.1 and the false alarm (FAR)/mis-detection (FRR) rates that result when these sensors are fused. A check mark in this table designates which of the sensors is included in the fusion process for that row. There are 1024 possible combinations. We selected these four to highlight the marginal contribution of stylometry and web browsing modalities when fused with the low level modalities. A closer inspection of the second and the fourth rows in the table reveals that discarding the stylometric sensors results in an order of magnitude deterioration in the global false alarm rate. This observation provides experimental support for the hypothesis that high level biometrics may be more beneficial as compared to low level modalities in the context of computer user detection accuracy. The plots in Figure 8.12 indicate that stylometry contributes more to reducing the error rates than web browsing.

8.7 Discussion and Future Work

We illustrated the use of behavioral sensors towards active authentication of users of computer systems. Four (4) behavioral features based on mouse and keyboard usage were used to train 4 classifiers which then became part of a distributed fusion scheme. Each classifier ("local detector") generated decisions on the hypothesis set, authenticating (or refusing to authenticate) a particular user, and the fusion center then performed a Neyman-Pearson test for a final authenticate/non-authenticate decision. ROC curves for the various fusion centers were generated, which revealed a hierarchy of sensor importance and would be helpful in identifying economical groups of sensors

to be used in user authentication. In future work, a wider variety of features could be included and an attempt might be made to tune the local detectors better toward improving the performance of the fusion center.

Even though it seems that stylometry is relatively the most efficient in identifying users, it has the inherent dis-advantage of requiring large amounts of data for proper classification. In general, more data would increase classification accuracy but compromise a quick detection. On the other hand, using small data collection windows would result in almost real time classification but with higher false alarm and mis-detection rates. This trade-off must be considered before employing stylometric sensors for real time user authentication.

Improved features sets from both low and high level modalities could provide a robust training sample for a particular user and therefore would lead to lower classification error rates when the validation system is used for identifying unknown users. The idea of fusing multiple modalities can be extended to identify adversarial users who might attempt partial spoofing (imitating authorized users) [120].

Furthermore, the idea of using behavioral features to develop unique user traits can also be applied for validating mobile device users. Features like texts, GPS location or applications used (frequency and type) can be extracted to train support vector machine or logistic regression based classifiers. Outputs from such classifiers can be fused to authenticate the identity of the device user.

9: HYPOXIA DETECTION USING KALMAN FILTER

In this chapter, we develop models and fusion rules for oximeters that detect the onset of hypoxia. Hypoxia is a medical condition affecting portions of the body that are deprived of oxygen supply. Prolonged exposure to cerebral oxygen deficiency can lead to unconsciousness or even death. The onset of hypoxia in humans is of concern for those operating in high altitudes, and in military flights characterized by high-acceleration maneuvers. Using oximeters for measuring blood oxygen saturation levels is a common means to detect hypoxia in real time. Many types of oximeters can be used for this task but all are prone to complicated noise characteristics and bias inaccuracies. It may therefore be advisable to collect and combine data streams from multiple oximeters for more reliable Hypoxia/No Hypoxia decisions (compared to decisions made by a single oximeter). Here we develop statistical noise models for three popular types of oximeters (Respironics Novametrix 515B, Nonin forehead pulse oximeter 9847, and Masimo Rad-87). We also combine data streams from these oximeters using a Kalman filter. The result is a smooth and reliable estimate of blood oxygen saturation level which can be used to detect the onset of hypoxia [126].

9.1 Context and Relevant Work

Hypoxia is diminished availability of oxygen to the cells of the body [127]. It can occur due to inadequate oxygenation of the lungs for extrinsic reasons, deficiency of oxygen in atmosphere, venous-to-arterial shunts (intrapulmonary or intra cardiac),

inadequate transport and delivery of oxygen, or inadequate tissue oxygenation or oxygen use. Exposure to severe hypoxia can lead to death of cells and depressed mental activity. Sometimes it culminates in coma and reduced work capacity of the muscles. Hypoxia occurs most commonly in people traveling to high altitude, performing strenuous exercise or work for prolonged periods of time at high altitudes. Another population at risk is combatants such as fighter pilots who undertake high G maneuvers.

Measuring the blood oxygen saturation (SpO2) is the most common and easiest way to instrumentally determine the presence of hypoxia. A healthy human has on average a SpO2 value of 95-100%. SpO2 values below 90% are considered low, and are taken as a possible indication of onset of hypoxia. The most common non-invasive device used to measure blood oxygen saturation levels is the pulse oximeter. The device uses a photo detector to measure the difference in the extinction curves of hemoglobin and oxygenated hemoglobin using light of different wavelengths [128], [129]. The common types of oximeters are applied either on the finger or on the forehead of the subject being monitored.

Hypoxia monitoring has been reported in several previous studies. A hypoxia detection and warning system was patented as a Aviation Hypoxia Monitor [130], which has a single pulse oximeter attached to the ear and provides a visual and audio signal if the blood level of a subject decreases significantly. The Hypoxia Detection and Warning System in [131] is composed of an electrochemical oxygen sensor located within the breathing mask of a pilot. It provides a vibratory warning within the mask when partial pressure of oxygen in the system falls below a set point. In [132],

a personal hypoxia monitoring system is proposed which uses the cross-correlation between heart rate, respiratory rate, blood flow velocity and blood oxygen saturation levels to identify the onset of hypoxia.

Even though pulse oximeters are very popular in operating rooms, emergency medical aids, and ambulatory use by heart and respiratory-system patients, oximeters are prone to inaccuracies due to several sources, most notably light scattering inside blood tissues. They are also affected by noise artifacts due to motion, ambient light interference, respiratory maneuvers, and pooling of blood at the point of measurement due to body orientation. In situations where fast and reliable hypoxia detection is required, a single pulse oximeter may not be sufficient, and it may be advantageous to use a combination of several such devices.

In a study conducted at the Naval Air Warfare Center Aircraft Division (NAWCAD) [133], three oximeters from different manufacturers were used simultaneously. These were Respironics Novametrix 515B (transmittance type on finger), and two reflectance type oximeters - Nonin pulse oximeter 9847 and Masimo Rad-87 (used on the forehead). The current study describes an attempt to fuse their observations using Kalman filtering so as to obtain a smoother and more reliable estimate of the blood oxygen saturation level than what one can get from a stand-alone single oximeter. The algorithm we propose can be executed in real time and has moderate computational requirements (computations can be carried out using wearable processors).

9.2 Data Collection

We used raw pulse oximeter data from the Time of Useful Consciousness study [133] carried out by NAWCAD, Patuxent River, MD. The study used 45 datasets from 26 volunteers (4 females and 22 males) who provided their informed consent under an approved NAWCAD IRB human research protocol. The subjects were exposed to a varying altitude profile ranging from 0 to 18,000 ft, simulated using a Reduced Oxygen Breathing Device (ROBD) [134]. The profile ascended at 1,000 ft/s to 10,000 ft and remained unchanged for 10 minutes, then ascended to 18,000 ft at the same rate and remained there for 20 minutes, and then descended at the same rate to ground level (0 ft). The volunteers spent up to 20 minutes at the equivalent of maximum altitude of 18,000 ft, during which time the data from a finger pulse oximeter (Respironics Novametrix 515B), and two forehead pulse oximeters (Nonin 9847 and Masimo RAD-87) were recorded. Subjects were exposed to one to three repetitions of the profiles.

9.3 Background

9.3.1 Kalman Filter

We use the standard model of a discrete dynamic system as a first order linear difference equation,

$$x(k) = F(k)x(k-1) + B(k)u(k) + G(k)w(k), \tag{9.1}$$

where $x(k) \in \mathbb{R}^n$ is the state of interest at time instant k, $u(k) \in \mathbb{R}^r$ is a known control input and $w(k) \in \mathbb{R}^q$ is a random vector referred to as the process noise. $F \in \mathbb{R}^{n \times n}$ is

the system matrix relating past state to the state at time instant k. $B \in \mathbb{R}^{n\times r}$ defines the influence of control inputs on the state at time k. The matrix $G \in \mathbb{R}^{n\times q}$ relates the process noise to the state at time k. The state observation model is defined as

$$z(k) = H(k)x(k) + v(k), \tag{9.2}$$

where $z(k) \in \mathbb{R}^m (m \leq n)$ is the observation vector, $H(k) \in \mathbb{R}^{m\times n}$ is the observation matrix at time instant k, and $v(k) \in \mathbb{R}^m$ is the measurement noise.

In this general setup, the optimal filtering problem is to estimate the state $x(k)$ at every time instant using only the noisy observations $z(k)$. When the estimator is assumed to be linear in the state variables, the standard Kalman Filter (KF) [12] is the best linear estimator in the mean squared error sense. The KF formulation below assumes that the process and measurement noises are uncorrelated, Gaussian and have zero mean, namely

$$E[w(k)] = E[v(k)] = 0, \quad \forall k, \tag{9.3}$$

with corresponding covariance

$$E[w(k)w^T(l)] = Q(k)\delta_{kl}, \tag{9.4}$$

$$E[v(k)v^T(l)] = R(k)\delta_{kl}. \tag{9.5}$$

Here $\delta()$ is Kronecker's delta. The process and measurement noises are also assumed

to have no cross correlation, namely

$$E[w(k)v^T(l)] = 0, \forall k, l. \tag{9.6}$$

When the Gaussian assumptions on the measurement and process noises are relaxed, the KF, though suboptimal, still remains the best linear estimator. A number of variants of KF were developed to deal with nonlinear systems (Extended KF, Unscented KF), and correlated noise (e.g., State Augmentation approach, measurement differencing). See [13, 14] for additional information about extensions of the KF.

The KF algorithm produces state estimates that minimize the mean-squared estimation error conditioned on a given observation sequence. The estimate of the state at a time k given all the information up to and including time k will be represented as $\hat{x}(k|k)$. The estimate of the state at time k given only information up to time $k-1$ is the one step prediction and is denoted as $\hat{x}(k|k-1)$. The corresponding estimation error covariances are denoted respectively as $P(k|k)$ and $P(k|k-1)$. Starting with the initial estimates $\hat{x}(0|0)$ and $P(0|0)$, the KF estimation algorithm constitutes the following *Prediction* and *Measurement Update* steps:

Prediction of state and variance at time k

$$\hat{x}(k|k-1) = F(k)\hat{x}(k-1|k-1) + B(k)u(k) \tag{9.7}$$

$$\begin{aligned} P(k|k-1) &= F(k)P(k-1|k-1)F^T(k) \\ &\quad +G(k)Q(k)G^T(k) \end{aligned} \tag{9.8}$$

The update of state estimate and variance at time k on the basis of predicted state and variance from previous step and the new observation $z(k)$ is given by

$$\begin{aligned} \hat{x}(k|k) &= \hat{x}(k|k-1) \\ &\quad +K(k)[z(k)-H(k)\hat{x}(k|k-1)] \qquad (9.9) \\ P(k|k) &= [I-K(k)H(k)]P(k|k-1). \\ &\quad [I-K(k)H(k)]^T + K(k)R(k)K(k)^T, \qquad (9.10) \end{aligned}$$

where the Kalman gain $K(k)$ is defined as

$$\begin{aligned} K(k) &= P(k|k-1)H^T(k). \\ &\quad [H(k)P(k|k-1)H^T(k)+R(k)]^{-1}. \qquad (9.11) \end{aligned}$$

The quantity $I(k) = [z(k) - H(k)\hat{x}(k|k-1)]$ is called the innovation sequence. When the assumptions about process and measurement noise statistical characteristics are satisfied, the autocorrelation of the innovations sequence has an impulse at zero lag (white noise).

9.4 Fusion of Oximeter Signals

Quick and reliable detection of the onset of hypoxia is of paramount importance for efficient delivery of medical assistance to hypoxia victims. The majority of the current work on hypoxia detection has been focused on developing better stand-alone oximeters for monitoring blood oxygen saturation levels. Biological signals are

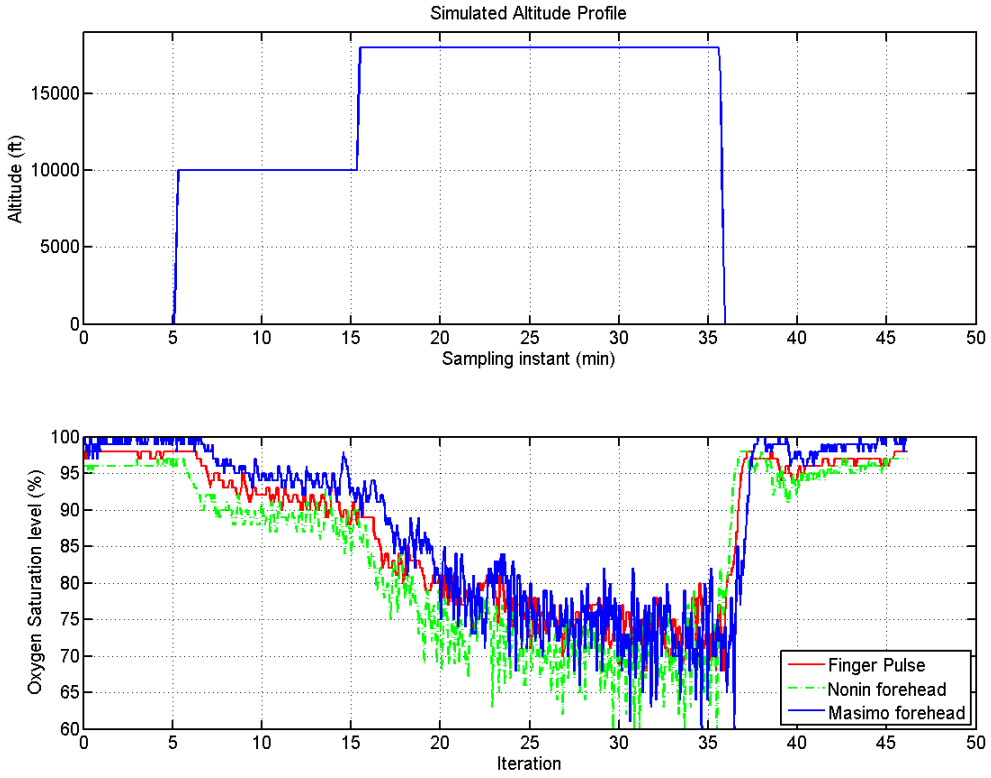


Figure 9.1: Simulated altitude profile and oximeter readings.

generally noisy and are subject to many unavoidable external factors such as motion and temperature fluctuations. Therefore, most oximeters when used independently in non-clinical dynamic settings tend to suffer from inaccuracies. The objective of the current study is to fuse several oximeter outputs in real time using an appropriately designed KF to generate an estimate of the blood oxygen saturation level which would be more accurate and reliable for hypoxia detection than any of the individual oximeter inputs. Figure 9.1 shows a sample of the collected data obtained from a particular subject. The top figure is the emulated altitude profile and the bottom figure is the raw observations (blood oxygen saturation levels in %) collected from the

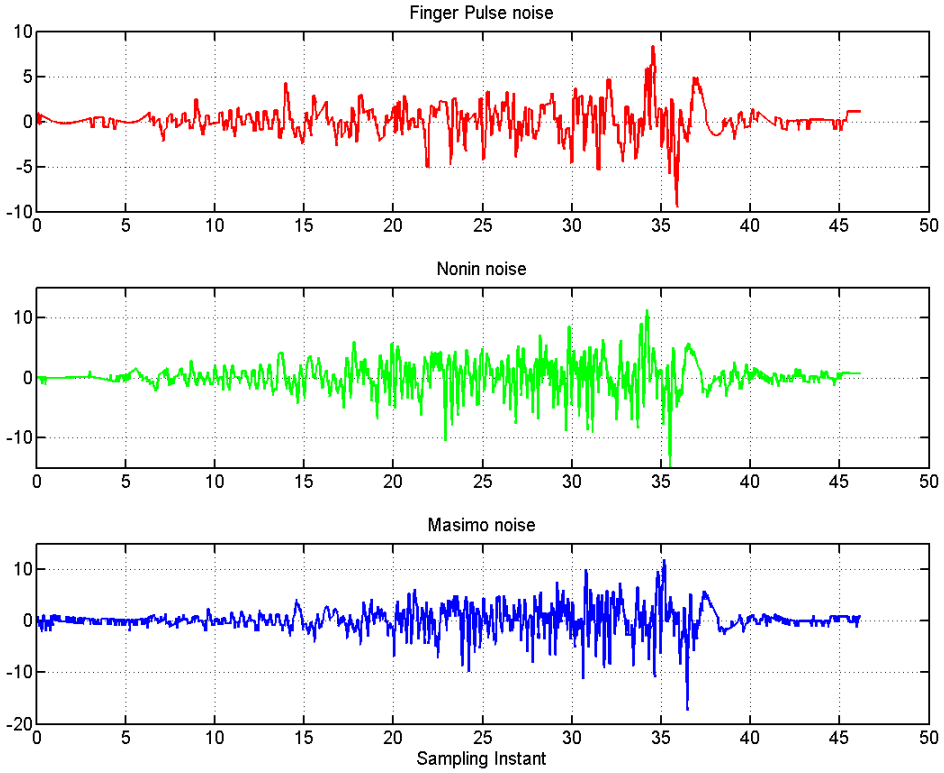


Figure 9.2: Oximeter noise for a particular subject.

three oximeters. The sampling rate of each oximeter was 60 Hz. The noise samples in these measurements were assumed to be high frequency and therefore extracted for each oximeter by low pass filtering the raw signals and collecting the residues. Figure 9.2 shows the extracted noise from the three oximeters for a particular subject. These noise samples are approximately zero mean but are temporally correlated. This property can be observed better through the normalized noise autocorrelations shown in Figure 9.3. Unlike uncorrelated noise (which has an autocorrelation function with an impulse at zero lag and negligible magnitude everywhere else), the oximeter noise samples we observed have wider decaying oscillatory autocorrelations. Moreover, the

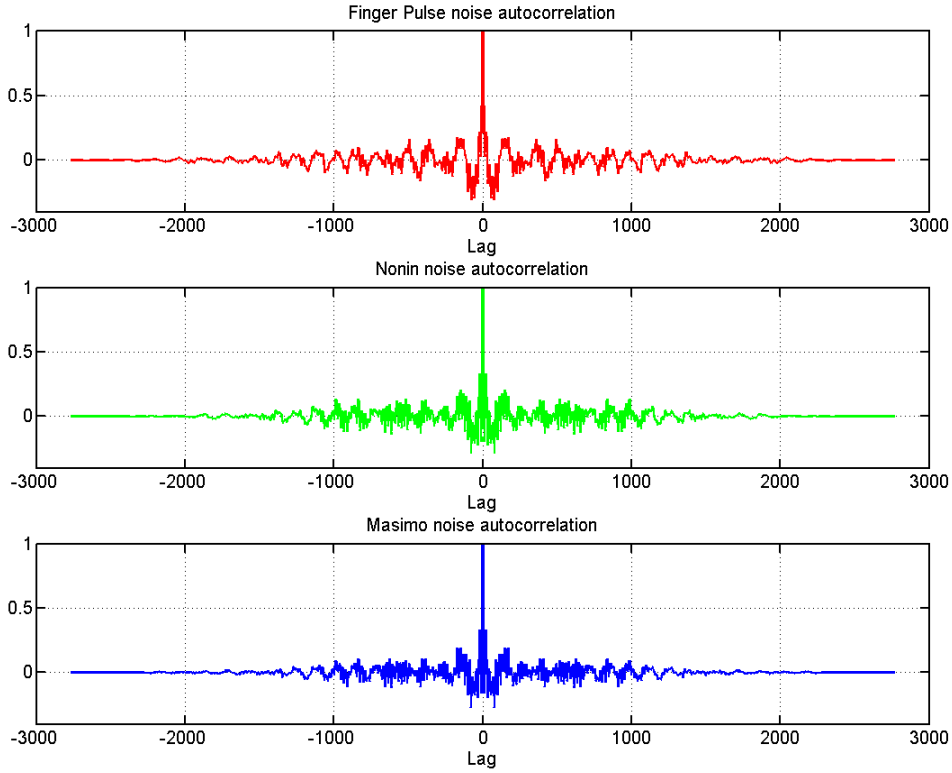


Figure 9.3: Oximeter noise autocorrelation.

statistics of the noise seem to change with time, which in our case correspond to the level of blood oxygen saturation (in the bottom figure of Figure 9.1, the oximeter readings tend to have higher variance when the signal level goes below 90%).

9.4.1 Oximeter Noise Model

One of the challenges in developing a proper KF for the the three oximeters is to incorporate the colored (temporally correlated) measurement noise observed in the oximeter measurement data in the KF's formulation and design. Here, we model the noise as the output of a second order Autoregressive (AR) process driven by zero mean white Gaussian noise. The decaying oscillations of the autocorrelation

of the noise samples (see Figure 9.3) motivated our choice of the order of the AR process (AR(2) process favors change in sign between terms of the process and also exhibits oscillations). Furthermore, we would use the state augmentation approach ([13], section 7.2) to incorporate colored noise in the KF formulation. For a l-order AR model for the colored measurement noise, using state augmentation would result in the dimension of the system increasing by $l * M$, where M is the number of sensors being fused. Therefore, higher order AR models would require KF iterations over large matrices which are computationally costly and may exhibit undesirable numerical instability. The need to avoid the increase in dimension of system matrices when higher order AR models are used was another motivation to use an AR(2) model for the colored measurement noise. Furthermore, simulations showed marginal or no improvement in terms of least square error, when higher order AR models (order higher than 2) were used. The AR(2) model for a colored observation signal $v(k)$ (measurement noise from an oximeter) is

$$v(k) = a_1 v(k-1) + a_2 v(k-2) + e(k), \tag{9.12}$$

where a_i for $i = 1, 2$ are the AR parameters and $e(k) \sim \mathcal{N}(0, \sigma_e^2(k))$ is zero mean Gaussian noise input to the AR system. We denote the variance of the colored sequence $v(k)$ at time k by $\sigma_v^2(k)$. Multiple runs of data were recorded for each subject for the same altitude profile during the data collection phase. For each subject, data from a single run were used as a training set to estimate model parameters.

The training set data were used to estimate the AR(2) parameters $a = [a_1, a_2]^T$ in

(9.12) for each subject by a least squares approach which can be represented through the Normal equations ([135], Chapter 8, page 225) as follows:

$$a = (A^T A)^{-1} A^T b \tag{9.13}$$

where

$$A = \begin{bmatrix} v(2) & v(1) \\ v(3) & v(2) \\ & \ddots \\ v(n-1) & v(n-2) \end{bmatrix} \tag{9.14}$$

and $b = [v(3)\ v(4) \cdots v(n)]^T$. This estimation scheme was repeated for all the subjects in the data collection process, therefore providing a parameter set associated with each test subject. The process of parameter estimation helps in computing subject dependent noise models but may not be suitable for subjects outside the test set.

As observed before from Figure 9.1, the oximeter noise standard deviations change (increase) as the observed readings go below 90% blood oxygen saturation level. We denoted the observation of an oximeter at time k by $z(k)$, and predicted the standard deviations of the measurement noise at every time instant using a quadratic regression model

$$\sigma_{pred}(k) = \beta_1 + \beta_2 d(k) + \beta_3 [d(k)]^2, \tag{9.15}$$

where the predicted standard deviation $\sigma_{pred}(k)$ at time instant k, is a function of

$d(k) = z(k) - 90$; the difference between $z(k)$ (noisy blood oxygen saturation level measured by an oximeter at time k) and 90% saturation level. The parameters $\beta = [\beta_1\ \beta_2\ \beta_3]^T$ are again estimated using the Normal equations. The noise samples for each oximeter were retrieved after low pass filtering the raw captured signal in the training set and collecting the residue. The mean of the noise samples were small enough and for simplicity sake were assumed to be zero. These noise samples were used as observations in the Normal equations (used to define the matrices A and b) - for estimating the parameters β. Let us represent the upper and lower bounds within which the measurement noise standard deviation varies, by $Smax$ and $Smin$, respectively. These bounds can be estimated from the predicted standard deviation $\sigma_{pred}(k)$ (defined in (9.15)) as

$$\begin{aligned} Smax &= max(\sigma_{pred}(k)), \\ Smin &= min(\sigma_{pred}(k)). \end{aligned}$$

The computed bounds were used to generate a first-order model for the variation of measurement noise standard deviation versus the difference between the raw observations $z(k)$ and 90% blood oxygen saturation level, as shown in Figure 9.4.

At a time instant k, the standard deviation of the measurement noise $\sigma_v(k)$ can be estimated as

$$\sigma_v(k) = \left[\frac{Smax - Smin}{-50}\right](z(k) - 90 - 10) + Smin \tag{9.16}$$

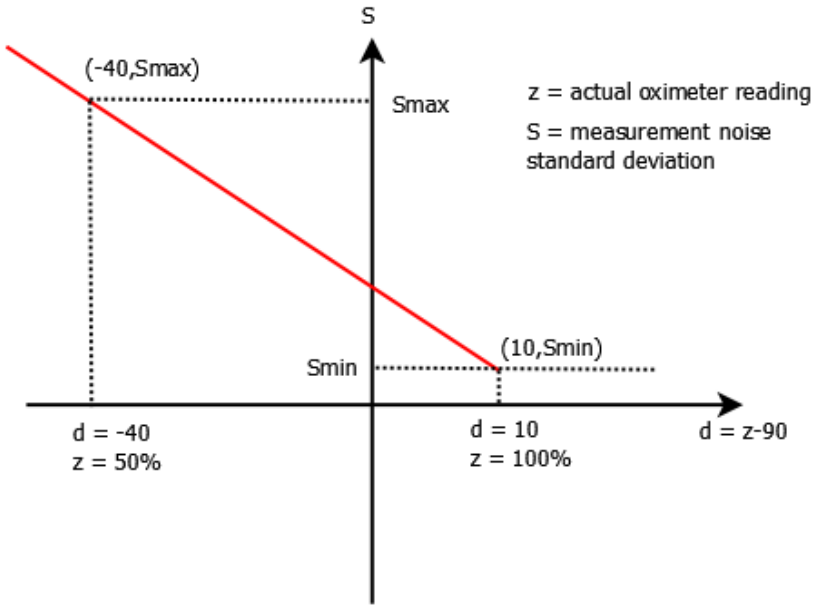


Figure 9.4: Oximeter noise variation model.

Next, let us denote the autocorrelation of the measurement noise $v(k)$ at lag j and at time instant k as $R^k_{vv}(j)$. The expressions for $R^k_{vv}(j)$ at lags 1 and 2 can be derived as (see [136], section 5.2 for details)

$$R^k_{vv}(1) = \left[\frac{a_1}{1-a_2}\right]\sigma^2_v(k), \text{and} \tag{9.17}$$

$$R^k_{vv}(2) = a_1 * R^k_{vv}(1) + a_2 * \sigma^2_v(k). \tag{9.18}$$

The variance of the zero mean Gaussian noise sequence $e(k)$ in (9.12) can then be estimated using (9.17) and (9.18) as

$$\sigma^2_e(k) = \sigma^2_v(k) - a_1 R^k_{vv}(1) - a_2 R^k_{vv}(2) \tag{9.19}$$

9.4.2 Kalman Filter Formulation

We aim to develop a KF formulation to estimate the blood oxygen saturation level by using data from 3 different oximeters. We consider the blood oxygen saturation level (in %) as the scalar state x to be estimated and define its dynamics with a first order system as

$$x(k) = Fx(k-1) + w(k), \tag{9.20}$$

where $x(k)$ is the actual blood oxygen saturation level at time instant k. The system matrix F is considered to be constant, equal to 1. $w(k)$ is the zero mean Gaussian process noise with variance Q. In other words, we model the evolution of the blood oxygen saturation level as a simple random walk. Since there are three (3) oximeters providing readings simultaneously at every time instant k, the measurement equation becomes

$$z(k) = \begin{bmatrix} z^1(k) \\ z^2(k) \\ z^3(k) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x(k) + \begin{bmatrix} v^1(k) \\ v^2(k) \\ v^3(k) \end{bmatrix} \tag{9.21}$$

where $z^i(k)$ and $v^i(k)$ are, respectively, the observation (noisy blood oxygen saturation level) and the measurement noise of the i^{th} oximeter for $i = 1, 2, 3$. We assume that the measurement noises are not correlated, i.e, $E[v^i(k)v^j(l)] = 0$, $\forall k, l$ and $i, j = 1, 2, 3, i \neq j$.

On the other hand, the measurement noise samples $v^i(k)$ are temporally correlated (colored). Colored measurement noise can be handled in multiple ways in the context

of Kalman filtering. Two of the most popular methods are the *state augmentation approach* ([13], section 7.2) and the *measurement differencing approach* ([137], section 11.2). Here we follow the standard state augmentation approach. It augments the the actual state vector $x(k)$ in (9.20) with the colored noise samples $v^i(k)$ which are the output of a linear system as defined in (9.12).

The AR(2) model in (9.12) has the state space representation

$$\begin{bmatrix} v(k) \\ v(k-1) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v(k-1) \\ v(k-2) \end{bmatrix} + \begin{bmatrix} e(k) \\ 0 \end{bmatrix} \tag{9.22}$$

The representation in (9.22) is used for modeling the measurement noises for all three (3) oximeters, but possibly with different parameters a_1 and a_2 for each oximeter. The original system in (9.20) is then augmented as shown in (9.23).

$$\begin{bmatrix} x(k) \\ v^1(k) \\ v^1(k-1) \\ v^2(k) \\ v^2(k-1) \\ v^3(k) \\ v^3(k-1) \end{bmatrix} = \begin{bmatrix} F & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_1^1 & a_2^1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_1^2 & a_2^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_1^3 & a_2^3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x(k-1) \\ v^1(k-1) \\ v^1(k-2) \\ v^2(k-1) \\ v^2(k-2) \\ v^3(k-1) \\ v^3(k-2) \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w(k) \\ e^1(k) \\ 0 \\ e^2(k) \\ 0 \\ e^3(k) \\ 0 \end{bmatrix} \tag{9.23}$$

In (9.23), $a_j^i, i = 1, 2, 3$ and $j = 1, 2$ are the AR parameters for the i^{th} oximeter estimated using (9.13). $e^i(k)$ is the input at time instant k for the AR model of the i^{th} oximeter measurement noise (see (9.12)). The corresponding augmented measurement

equation becomes

$$z(k) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ v^1(k) \\ v^1(k-1) \\ v^2(k) \\ v^2(k-1) \\ v^3(k) \\ v^3(k-1) \end{bmatrix} \tag{9.24}$$

The augmented measurement equation in (9.24) no longer has any direct measurement noise, and therefore the matrix R in (9.10) and (9.11) is 0. This can lead to numerical instability in the KF updating process. A diagonal matrix with a small trace can be used for R in order to ensure stability and non-singularity of the various matrices involved [138].

$$Q^a(k) = \begin{bmatrix} Q & & & & & & \\ & \sigma^2_{e^1}(k) & & & & & \\ & & 0 & & & 0 & \\ & & & \sigma^2_{e^2}(k) & & & \\ & & & & 0 & & \\ & 0 & & & & \sigma^2_{e^3}(k) & \\ & & & & & & 0 \end{bmatrix} \tag{9.25}$$

The standard KF, as described by equations (9.7) - (9.11), can then be applied on the augmented system to obtain the desired state estimate. The covariance matrix for the augmented process noise vector at time instant k represented as $Q^a(k)$ is given

in (9.25), where $\sigma^2_{e^i}(k), i = 1, 2, 3$ is the variance of the AR model input noise process for the i^{th} oximeter as computed from (9.19).

To make sure that the augmented covariance matrix is positive definite, the zero diagonal elements of $Q^a(k)$ can be replaced by a small positive number $\epsilon > 0$. This modification does not make appreciable difference to state dynamics since the corresponding rows in the process noise coefficient matrix in (9.23) are zero.

9.5 Model Validation with Synthetic Data

Analysis of the oximeter observations collected from the training set, showed that the measurement noise for each oximeter is temporally correlated and the variance also varies depending on whether the observation value is above or below 90% blood oxygen saturation level. This situation motivated us to use a second order AR process to model the measurement noise which was then incorporated into the KF using the state augmentation approach. To test the validity of the assumed noise models, synthetic blood oxygen saturation level data were generated and corrupted with temporally correlated noise. The standard deviation of the noise was made to vary as a first order function of the difference between the true data and 90%.

Observations for three sensors were simulated with colored measurement noise. Figure 9.5 shows the performance of the state augmented Kalman filter model when applied on the simulated data. The top figure shows the true signal (simulated blood oxygen saturation level), the noisy observations from the three (3) oximeters (true signal corrupted by temporally correlated noise with data dependent variance), and the KF estimate of the simulated blood oxygen saturation level. The bottom figure

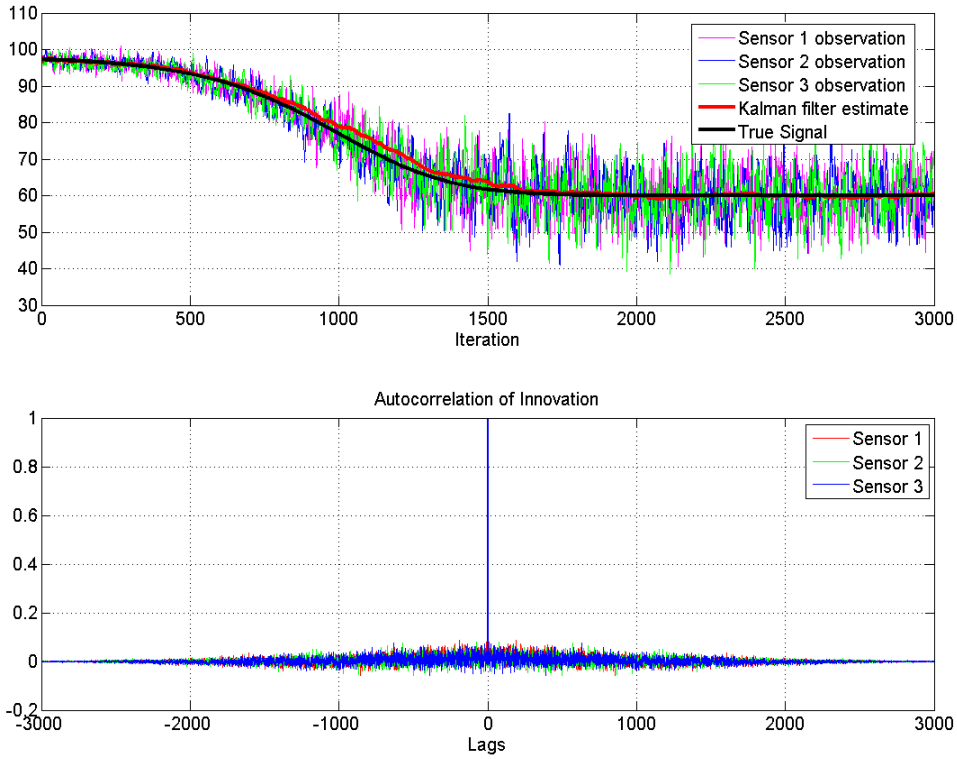


Figure 9.5: Performance of proposed Kalman filter model for simulated data.

shows the autocorrelation of the innovation sequences. Second order whiteness of a random sequence can be tested by a statistical test as defined in ([139], page 16-8). The test involves using the biased estimate of autocorrelation defined as

$$c_y(\tau) = \frac{1}{N}\sum_{t=1}^{N-t} y(t+\tau)y(t), \tau \geq 0 \tag{9.26}$$

to form the test statistic

$$T = \frac{N}{c_y^2(0)}\sum_{i=1}^{m} c_y^2(i). \tag{9.27}$$

Here the sequence y is zero mean with length N. The parameter m is a chosen maximum lag for whiteness testing. The test statistic T is distributed chi-squared, $\chi^2(m)$ (χ^2 with m degrees of freedom) if the sequence y is zero mean white. For a fixed significance level, the test statistic can be compared to a threshold to validate the second order whiteness. If T is greater than the threshold, the sequence y is declared non-white at the chosen significance level.

The test described above, when applied on the innovation sequence obtained after Kalman filtering the simulated data, declared the innovation sequences to be second order white at 0.05 significance level. The maximum lag m was chosen to be 100. The whiteness of the innovations statistically validates the functioning of the KF with the proposed noise models on simulated observations corrupted with colored measurement noise with time varying variance.

9.6 Filter Performance on Real Data

A standard KF was implemented on the augmented system defined by (9.23) and (9.24) with original process noise variance $Q = 0.005$. The AR model parameters and noise variances were estimated using (9.13) and (9.19) where the matrix A was built using the noise samples extracted after low pass filtering of the training data. For each subject, the estimation process was repeated for each oximeter to generate three(3) sets of parameters for the 3 corresponding oximeters. Figure. 9.6 shows the Kalman filtered estimate of the blood oxygen saturation level (in black line) given the raw observations from the three oximeters for a particular subject. The process noise variance Q was assumed to be a moderate 0.005, reflecting the average confidence on

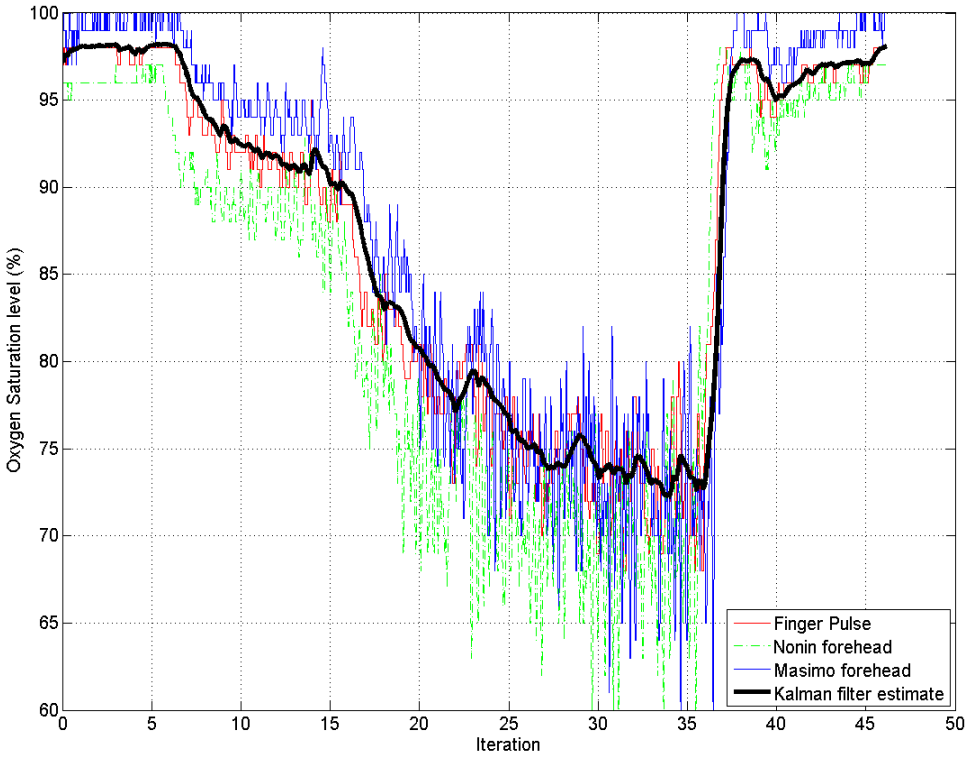


Figure 9.6: Kalman filter estimate and raw oximeter readings for blood oxygen saturation level.

the state dynamic model.

Since the random walk model in (9.20) is one of the first attempts to capture the dynamics of the blood oxygen saturation level, the choice of a moderate process noise variance is probably prudent.

The performance of the filter can be assessed by analyzing the autocorrelation of the innovation sequences. The top figure in Figure 9.7 shows the autocorrelation of the innovations obtained by applying the Kalman filter on simulated data (recall that the innovations from simulated data were statistically declared white). The middle figure provides the autocorrelation of the innovation sequences obtained after

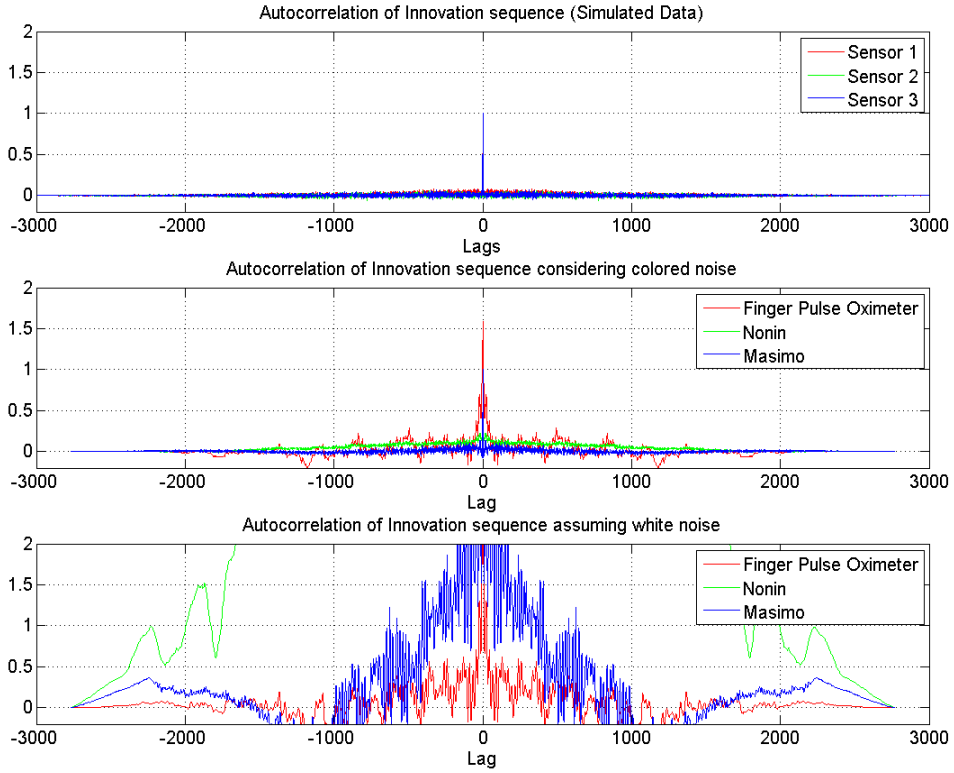


Figure 9.7: Autocorrelation of innovation sequences for three sensors (Real and Simulated).

Kalman filtering of the real data using the proposed colored noise models. The bottom figure shows the autocorrelation of the innovation sequences when KF was applied on the real data with the measurement noised assumed to be white. The colored noise model appear to significantly improve the filter performance. Even though the innovation sequences generated from the real data are not white, there is significantly less correlation among innovation samples at higher lags when compared to the autocorrelations shown in Figure 9.3.

9.7 Discussion and Future Work

We proposed the fusion of multiple oximeter signals using a Kalman filter where the colored measurement noise was modeled using a second order AR process. The results are not perfect (autocorrelation of the innovations not white); it would probably be useful to try other models of correlated noise. For instance a multiplicative noise model like the Autoregressive Conditional Heteroskedasticity (ARCH) [140], [141] can be attempted to model the oximeter measurement noise. Time varying autoregressive models (TVAR) may also be used. Other techniques like particle filtering ([13], Chapter 15) might be used to take into account non-Gaussian noise distributions.

The model parameters were derived using training data sets for each subject. In other words, for each subject tested in the data collection process, an individual set of parameters was estimated. These parameters however, may not be suitable for subjects not included in the data collection process. An on-line learning algorithm capable of reliably estimating model parameters *on the run* using incoming data from multiple sensors would be more useful as part of a generic real time hypoxia detection scheme. To enhance the design of such generic systems, future work could include improving the noise model, incorporating multi modal sensing, and developing pre-filtering steps based on human dynamics.

Our study assumed that the state dynamics are autonomous (zero input). However an attempt can be made to derive a relation between altitude and blood oxygen saturation level which could then possibly be incorporated in the KF formulation with the altitude variation being the input. This effort may also pave the way to incorporating

information from multiple heterogeneous sensors like altimeter, accelerometers, anti-G-suit pressure (for fighter pilots), in a combined framework which can potentially detect the onset of hypoxia more reliably in real time. The end combined result can be further used as an activation trigger signal to control hypoxia mitigation technologies as well.

In addition, Figures 9.2 and 9.3 show a strong similarity in noise characteristics for the three oximeters. This similarity might be due to some physiological processes in the body and perhaps is not a property of the oximeters. This observation needs to be investigated further.

10: SUMMARY

A multitude of issues under the context of distributed detection and estimation were presented in this thesis. Both *Hard* and *Hard/Soft* fusion scenarios were investigated and multiple schemes for tractable fusion were proposed. In hard sensor fusion, we proposed a new alternative algorithm that computes the fusion center and local detector operating points for a binary hypothesis decision fusion problem. Unlike the more traditional PBPO approach, the proposed algorithm does not require solutions of nonlinear coupled algebraic equations and is also guaranteed to achieve global optimality. A number of simulated example cases demonstrated the degree by which the proposed algorithm outperformed the traditional approach.

In *Hard/Soft* sensor fusion, we considered the special situation in distributed detection scheme where apart from hard sensors, humans also act as information sources. The literature at present lacks well defined hard/soft fusion algorithms and there is even more scarcity of practical examples. The hard/soft fusion research discussed in this thesis attempted to alleviate this scarcity of well defined hard/soft fusion schemes and presented practical and implementable frameworks and algorithms. Techniques from evidence theory and belief calculus were used to develop a versatile hard/soft fusion framework. The proposed scheme was tested against a variety of simulated example scenarios to demonstrate the range of applicability of the algorithm. Furthermore, an original tree based hierarchical structure was developed to represent a

class of tentative and equivocal human opinions called *Conditionally Refined opinion (CRo)*. Rigorous mathematical algorithms for belief propagation across the levels of hierarchy were proposed as well. Several algorithms were developed for both soft/soft and hard/soft fusion of information represented using a CRo.

The final part of the thesis discussed applications of distributed detection techniques on two real world problems. We employed parallel binary decision fusion algorithms to differentiate between legitimate and unauthorized computer users. A wide variety of biometric sensors which monitored behavioral characteristics of the user were chosen and formed the bank of local detectors. ROC curves showed the advantage of fusion through the increased detection performance as compared to individual sensors/detectors. In the second case, we used state augmented Kalman filter to estimate hypoxia by measuring blood oxygen saturation levels in various individuals exposed to high altitudes. In the process, statistical pulse oximeter noise models were developed. The study showed significant improvements in the estimated saturation levels and possibly could be used as the ground work to validate and build a robust real time hypoxia detection system.

Bibliography

[1] Belur V Dasarathy. *Decision Fusion*, volume 1994. IEEE Computer Society Press Los Alamitos, CA, 1994.

[2] David L Hall, Michael McNeese, James Llinas, and Tracy Mullen. A Framework for Dynamic Hard/Soft Fusion. In *11th International Conference on Information Fusion*, pages 1–8, 2008.

[3] R.R. Tenney and N.R. Sandell. Detection with Distributed Sensors. *IEEE Transactions on Aerospace and Electronic Systems*, AES-17(4):501 –510, July 1981.

[4] Hugh Durrant-Whyte and Thomas C Henderson. Multisensor Data Fusion. *Springer handbook of robotics*, pages 585–610, 2008.

[5] Samawal Alhakeem and Pramod K Varshney. Decentralized Bayesian Detection with Feedback. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 26(4):503–513, 1996.

[6] Moshe Kam, Chris Rorres, Wei Chang, and Xiaoxun Zhu. Performance and Geometric Interpretation for Decision Fusion with Memory. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,*, 29(1):52–62, 1999.

[7] Dimitrios A Pados, Karen W Halford, Dimitri Kazakos, and Panayota Papantoni-Kazakos. Distributed Binary Hypothesis Testing with Feedback. *IEEE Transactions on Systems, Man and Cybernetics*, 25(1):21–42, 1995.

[8] Peter F Swaszek and Peter Willett. Parley as an Approach to Distributed Detection. *Aerospace and Electronic Systems, IEEE Transactions on*, 31(1):447–457, 1995.

[9] Ramanarayanan Viswanathan and Pramod K Varshney. *Distributed Detection with Multiple Sensors I. Fundamentals*, 1997.

[10] Jason D Papastavrou. *Decentralized Decision Making in a Hypothesis Testing Environment*. PhD thesis, Massachusetts Institute of Technology, 1990.

[11] Pramod K Varshney. *Distributed Detection and Data Fusion*. Springer-Verlag New York, Inc., 1996.

[12] Rudolph Emil Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[13] Dan Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.

[14] A.P. Sage and J.L. Melsa. *Estimation Theory With Applications To Communications And Control*. McGraw-Hill series in Systems Science. McGraw-Hill, 1974.

[15] Hugh F Durrant-Whyte. Introduction to Estimation and the Kalman Filter. *Australian Centre for Field Robotics*, 2001.

[16] Glenn Shafer. *A Mathematical Theory of Evidence*, volume 1. Princeton University Press, Princeton, 1976.

[17] J. Tsitsiklis and M. Athans. On the Complexity of Decentralized Decision Making and Detection Problems. *IEEE Transactions on Automatic Control*, AC-30(5):440–446, 1985.

[18] F.A. Sadjadi. Hypotheses Testing in a Distributed Environment. *IEEE Transactions on Aerospace and Electronic Systems*, AES-22(2):134 –137, March 1986.

[19] H.L. Van Trees. *Detection, Estimation, and Modulation Theory*. Number pt. 1 in Detection, Estimation, and Modulation Theory. Wiley, 2004.

[20] Z. Chair and P.K. Varshney. Optimal Data Fusion in Multiple Sensor Detection Systems. *IEEE Transactions on Aerospace and Electronic Systems*, AES-22(1):98 –101, Jan. 1986.

[21] Z Chair and Pramod K Varshney. Distributed Bayesian Hypothesis Testing with Distributed Data Fusion. *IEEE Transactions on Systems, Man and Cybernetics,*, 18(5):695–699, 1988.

[22] A.R. Reibman and L.W. Nolte. Optimal Detection and Performance of Distributed Sensor Systems. *IEEE Transactions on Aerospace and Electronic Systems*, AES-23(1):24 –30, Jan. 1987.

[23] M. Kam, Q. Zhu, and W.S. Gray. Optimal Data Fusion of Correlated Local Decisions in Multiple Sensor Detection Systems. *IEEE Transactions on Aerospace and Electronic Systems*, AES-28(3):916 –920, Jul 1992.

[24] V. Aalo and R. Viswanathan. On Distributed Detection with Correlated Sensors: Two Examples. *IEEE Transactions on Aerospace and Electronic Systems*, AES-25(3):414 –421, May 1989.

[25] E Drakopoulos and C C Lee. Optimum Fusion of Correlated Local Decisions. In *Proceedings of the 27th IEEE Conference on Decision and Control,*, pages 2489–2494. IEEE, 1988.

[26] Roman Krzysztofowicz and Dou Long. Fusion of detection probabilities and comparison of multisensor systems. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):665–677, 1990.

[27] Dimitris A Pados, P Papantoni-Kazakos, Demetrios Kazakos, and Achilles G Koyiantis. On-line Threshold Learning for Neyman-Pearson Distributed Detection. *IEEE Transactions on Systems, Man and Cybernetics*, 24(10):1519–1531, 1994.

[28] S.C.A. Thomopoulos, R. Viswanathan, and D.K. Bougoulias. Optimal Distributed Decision Fusion. *IEEE Transactions on Aerospace and Electronic Systems*, AES-25(5):761 –765, Sep 1989.

[29] Imad Y Hoballah and Pramod K Varshney. An Information Theoretic Approach to the Distributed Detection Problem. *IEEE Transactions on Information Theory*, 35(5):988–994, 1989.

[30] I.Y. Hoballah and P.K. Varshney. Distributed Bayesian Signal Detection. *IEEE Transactions on Information Theory*, IT-35(5):995 –1000, Sep 1989.

[31] Imad Y. Hoballah and P. K. Varshney. Neyman-Pearson Detection with Distributed Sensors. In *25th IEEE Conference on Decision and Control*, volume 25, pages 237 –241, Dec 1986.

[32] I. Y. Hoballah. *On the Design and Optimization of Distributed Signal Detection and Parameter Estimation Systems,*. PhD thesis, Syracuse Univ., Syracuse, NY, Dec. 1986.

[33] John N Tsitsiklis. On Threshold Rules in Decentralized Detection. In *25th IEEE Conference on Decision and Control,*, volume 25, pages 232–236. IEEE, 1986.

[34] William W Irving and John N Tsitsiklis. Some Properties of Optimal Thresholds in Decentralized Detection. *IEEE Transactions on Automatic Control,*, 39(4):835–838, 1994.

[35] Po-Ning Chen and A. Papamarcou. New Asymptotic Results in Parallel Distributed Detection. *IEEE Transactions on Information Theory*, IT-39(6):1847–1863, 1993.

[36] John N Tsitsiklis. Decentralized Detection by a Large Number of Sensors. *Mathematics of Control, Signals and Systems*, 1(2):167–182, 1988.

[37] Dario Bauso and R. Pesenti. Generalized Person-By-Person Optimization in Team Problems with Binary Decisions. In *American Control Conference*, pages 717–722, 2008.

[38] Sayandep Acharya, Ji Wang, and Moshe Kam. Distributed Decision Fusion using the Neyman-Pearson Criterion. In *17th International Conference on Information Fusion (FUSION)*, In Press.

[39] John N. Tsitsiklis. Decentralized Detection. In *In Advances in Statistical Signal Processing*, pages 297–344, 1993. JAI Press.

[40] Victor Y Pan. Univariate Polynomials: Nearly Optimal Algorithms for Numerical Factorization and Root-finding. *Journal of Symbolic Computation*, 33(5):701–733, 2002.

[41] Markus Lang and B-C Frenzel. Polynomial Root Finding. *Signal Processing Letters, IEEE*, 1(10):141–143, 1994.

[42] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):pp. 325–339, 1967.

[43] Zdzisław Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*, volume 9. Springer, 1991.

[44] Lotfi A Zadeh. Fuzzy Sets. *Information and Control*, 8(3):338–353, 1965.

[45] Didier Dubois, Henri M Prade, Henri Farreny, Roger Martin-Clouaire, and Claudette Testemale. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, volume 2. Plenum press New York, 1988.

[46] Didier Dubois and Henri Prade. Possibility Theory in Information Fusion. In *Proceedings of the Third International Conference on Information Fusion (FUSION)*, volume 1, pages PS6–P19. IEEE, 2000.

[47] Ronald R Yager, Liping Liu, et al. *Classic Works of The Dempster-Shafer Theory of Belief Functions*, volume 219. Springer, 2008.

[48] Enrique H Ruspini, John D Lowrance, and Thomas M Strat. Understanding Evidential Reasoning. *International Journal of Approximate Reasoning*, 6(3):401–424, 1992.

[49] Philippe Smets. Data Fusion in The Transferable Belief Model. In *Proceedings of the Third International Conference on Information Fusion (FUSION)*, volume 1, pages PS21–PS33. IEEE, 2000.

[50] Jürg Kohlas and Paul-André Monney. Theory of Evidencea Survey of its Mathematical Foundations, Applications and Computational Aspects. *Zeitschrift für Operations Research*, 39(1):35–68, 1994.

[51] Philippe Smets and Robert Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66(2):191–234, 1994.

[52] Philippe Smets. The Combination of Evidence in The Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.

[53] Glenn Shafer and Judea Pearl. *Readings in Uncertain Reasoning*. Morgan Kaufmann Publishers Inc., 1990.

[54] Andrej Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, volume 3. Berlin, 1933.

[55] Lotfi A Zadeh. *On the Validity of Dempster's Rule of Combination of Evidence*. Electronics Research Laboratory, University of California, 1979.

[56] Lotfi A Zadeh. Review of A Mathematical Theory of Evidence. *AI magazine*, 5(3):81, 1984.

[57] Kari Sentz and Scott Ferson. *Combination of Evidence in Dempster-Shafer Theory*, volume 4015. Citeseer, 2002.

[58] Ronald R Yager. On the Dempster-Shafer Framework and New Combination Rules. *Information Sciences*, 41(2):93–137, 1987.

[59] Audun Jøsang, Milan Daniel, and Patrick Vannoorenberghe. Strategies for Combining Conflicting Dogmatic Beliefs. In *Proceedings of the 6th International Conference on Information Fusion*, pages 1133–1140, 2003.

[60] Audun Jøsang. The Consensus Operator for Combining Beliefs. *Artificial Intelligence*, 141(1):157–170, 2002.

[61] Audun Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001.

[62] Florentin Smarandache and Jean Dezert. *Advances and Applications of DSmT for Information Fusion (Collected works), second volume: Collected Works*, volume 2. Infinite Study, 2006.

[63] Florentin Smarandache and Jean Dezert. Information Fusion Based on New Proportional Conflict Redistribution Rules. In *2005 8th International Conference on Information Fusion*, volume 2, pages 8–pp. IEEE, 2005.

[64] Didier Dubois and Henri Prade. A Set-Theoretic View of Belief Functions Logical Operations and Approximations by Fuzzy Sets. *International Journal Of General System*, 12(3):193–226, 1986.

[65] Didier Dubois and Henri Prade. On the Combination of Evidence in Various Mathematical Frameworks. In *Reliability Data Collection And Analysis*, pages 213–241. Springer, 1992.

[66] Philippe Smets. Decision Making in The tbm: The Necessity of The Pignistic Transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.

[67] Philippe Smets. Decision Making in a Context Where Uncertainty is Represented by Belief Functions. In *Belief functions in business decisions*, pages 17–61. Springer, 2002.

[68] Fabio Cuzzolin. A Geometric Approach to The Theory of Evidence. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(4):522–534, 2008.

[69] John J Sudano and Lockheed Martin. Yet Another Paradigm Illustrating Evidence Fusion. In *2006 9th International Conference on Information Fusion*, pages 1–7. IEEE, 2006.

[70] Jean Dezert and Florentin Smarandache. A New Probabilistic Transformation of Belief Mass Assignment. In *2008 11th International Conference on Information Fusion*, pages 1–8. IEEE, 2008.

[71] Pekka Orponen. Dempster's Rule of Combination is P-complete. *Artificial Intelligence*, 44(1):245–253, 1990.

[72] D.L. Hall and J. Llinas. *Multisensor Data Fusion, -2 Volume Set*. CRC, 2001.

[73] James Llinas, Rakesh Nagi, David Hall, and John Lavery. A Multi-Disciplinary University Research Initiative in Hard and Soft Information Fusion: Overview, Research Strategies and Initial Results. In *13th International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE, 2010.

[74] TL Wickramarathne, Kamal Premaratne, Manohar N Murthi, Matthias Scheutz, S Kubler, and M Pravia. Belief Theoretic Methods for Soft and Hard Data Fusion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2388–2391. IEEE, 2011.

[75] Jacob L Graham, David L Hall, and Jeffrey Rimland. A coin-inspired synthetic dataset for qualitative evaluation of hard and soft fusion systems. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2011.

[76] Jeffrey C Rimland, Dan Coughlin, David L Hall, and Jacob L Graham. Advances in Data Representation for Hard/Soft Information Fusion. In *SPIE Defense, Security, and Sensing*, pages 84070Q–84070Q. International Society for Optics and Photonics, 2012.

[77] Sayandeep Acharya and Moshe Kam. Evidence Combination for Hard and Soft Sensor Data Fusion. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2011.

[78] Audun Jøsang, Simon Pope, and David McAnally. Normalising the Consensus Operator for Belief Fusion. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty (IPMU2006)*, 2006.

[79] Face recognition based on statistical moments, Available: http://www.advancedsourcecode.com/statisticalmoments.asp.

[80] R.C. Gonzalez, R.E. Woods, and S.L. Eddins. *Digital Image Processing Using MATLAB*. Pearson Education, 2004.

[81] Jean Gordon and Edward H Shortliffe. A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space. *Artificial Intelligence*, 26(3):323–357, 1985.

[82] Glenn Shafer and Roger Logan. Implementing Dempster's Rule for Hierarchical Evidence. *Artificial Intelligence*, 33(3):271–298, 1987.

[83] Judea Pearl. On Evidential Reasoning in a Hierarchy of Hypotheses. *Artificial Intelligence*, 28(1):9–15, 1986.

[84] Judea Pearl. Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, 29(3):241–288, 1986.

[85] Glenn Shafer, Prakash P Shenoy, and Khaled Mellouli. Propagating Belief Functions in Qualitative Markov Trees. *International Journal of Approximate Reasoning*, 1(4):349–400, 1987.

[86] Arthur P Dempster and Augustine Kong. Uncertain Evidence and Artificial Analysis. *Journal of Statistical Planning and Inference*, 20(3):355–368, 1988.

[87] Glenn Shafer. Hierarchical evidence. In *CAIA*, pages 16–21, 1985.

[88] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[89] Timothy J Pleskac and Jerome R Busemeyer. Two-stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence. *Psychological Review*, 117(3):864, 2010.

[90] Donald J. Bucci, Sayandeep. Acharya, and Moshe Kam. Simulating Human Decision Making for Testing Soft and Hard/Soft Fusion Algorithms. In *47th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2013.

[91] Donald J Bucci, Sayandeep Acharya, Timothy J Pleskac, and Moshe Kam. Subjective Confidence and Source Reliability in Soft Data Fusion. In *48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014.

[92] Sayandeep Acharya, Donald J Bucci, and Moshe Kam. Representation and Fusion of Conditionally Refined Opinions Using Evidence Trees. In *15th International Conference on Information Fusion (FUSION)*, pages 939–946. IEEE, 2012.

[93] Y. Sui, X. Zou, E.Y. Du, and F. Li. Secure and privacy-preserving biometrics based active authentication. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1291–1296. IEEE, 2012.

[94] S.P. Banerjee and D.L. Woodard. Biometric Authentication and Identification using Keystroke Dynamics: A Survey. *Journal of Pattern Recognition Research*, 7:116–139, 2012.

[95] A. Ross and A.K. Jain. Multimodal Biometrics: An Overview. In *Proceedings of 12th European Signal Processing Conference*, pages 1221–1224, 2004.

[96] J. Cui, J.P. Li, and X.J. Lu. Study on Multi-Biometric Feature Fusion and Recognition Model. In *International Conference on Apperceiving Computing and Intelligence Analysis (ICACIA)*, pages 66–69. IEEE, 2008.

[97] S.C.A. Thomopoulos, R. Viswanathan, and D.C. Bougoulias. Optimal decision fusion in multiple sensor systems. *IEEE Transactions on Aerospace and Electronic Systems,*, AES-23(5):644 –653, sept. 1987.

[98] S. Bleha, C. Slivinsky, and B. Hussien. Computer-Access Security Systems Using Keystroke Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(12):1217–1222, 1990.

[99] K. STUART, K. William, and J. BETTY. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a crt. *Ergonomics*, 21(8):601–613, 1978.

[100] M. Karnan, M. Akila, and N. Krishnaraj. Biometric Personal Authentication using Keystroke Dynamics: A Review. *Applied Soft Computing*, 11(2):1565–1573, 2011.

[101] N. Bartlow and B. Cukic. Evaluating the Reliability of Credential Hardening through Keystroke Dynamics. In *International Symposium on Software Reliability Engineering (ISSRE)*, pages 117–126. IEEE, 2006.

[102] R. Giot, M. El-Abed, and C. Rosenberger. Keystroke Dynamics Authentication for Collaborative Systems. In *International Symposium on Collaborative Technologies and Systems (CTS)*, pages 172–179. IEEE, 2009.

[103] F. Monrose and A.D. Rubin. Keystroke Dynamics as a Biometric for Authentication. *Future Generation Computer Systems*, 16(4):351–359, 2000.

[104] M. Choraś and P. Mroczkowski. Keystroke Dynamics for Biometrics Identification. *Adaptive and Natural Computing Algorithms*, pages 424–431, 2007.

[105] P. Mroczkowski. *Identity Verification using Keyboard Statistics.* PhD thesis, Linköping, 2004.

[106] M. Pusara and C.E. Brodley. User Re-Authentication via Mouse Movements. In *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*, pages 1–8. ACM, 2004.

[107] A.A.E. Ahmed and I. Traore. A New Biometric Technology Based on Mouse Dynamics. *IEEE Transactions on Dependable and Secure Computing*, 4(3):165 –179, july-sept. 2007.

[108] Nan Zheng, Aaron Paloski, and Haining Wang. An Efficient User Verification System via Mouse Movements. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*, CCS '11, pages 139–150, New York, NY, USA, 2011. ACM.

[109] Z. Jorgensen and T. Yu. On Mouse Dynamics as a Behavioral Biometric for Authentication. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, pages 476–482. ACM, 2011.

[110] D. Umphress and G. Williams. Identity Verification Through Keyboard Characteristics. *International Journal of Man-Machine Studies*, 23(3):263–273, 1985.

[111] N.V. Boulgouris, K.N. Plataniotis, and E. Micheli-Tzanakou. *Biometrics: Theory, Methods, and Applications*, volume 9. Wiley-IEEE Press, 2009.

[112] S.A Bleha, J. Knopp, and M.S Obaidat. Performance of the Perceptron Algorithm for the Classification of Computer Users. In *Proceedings of the 1992 ACM/SIGAPP Symposium on Applied computing: Technological Challenges of the 1990's*, pages 863–866. ACM, 1992.

[113] Q. Tao and R. Veldhuis. Robust Biometric Score Fusion by Naive Likelihood Ratio via Receiver Operating Characteristics. *IEEE Transactions on Information Forensics and Security,*, PP(99):1, 2012.

[114] S.K. Sahoo, T. Choubisa, et al. Multimodal Biometric Person Authentication: A Review. *IETE Technical Review*, 29(1):54, 2012.

[115] C. Sanderson and K.K. Paliwal. Identity Verification Using Speech and Face Information. *Digital Signal Processing*, 14(5):449–480, 2004.

[116] A. Jordan. On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes. *Advances in Neural Information Processing Systems*, 14:841, 2002.

[117] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory*. Wiley, 2004.

[118] Alex Fridman, Ariel Stolerman, Sayandeep Acharya, Patrick Brennan, Patrick Juola, Rachel Greenstadt, and Moshe Kam. Active Authentication Linguistic Modalities. Technical report, DTIC Document, 2013.

[119] A. Fridman, A. Stolerman, S. Acharya, P. Brennan, P. Juola, R. Greenstadt, and M. Kam. Decision Fusion for Multimodal Active Authentication. *IT Professional*, 15(4):29–33, July 2013.

[120] Alex Fridman, Ariel Stolerman, Sayandeep Acharya, Patrick Brennan, Patrick Juola, Rachel Greenstadt, and Moshe Kam. Multi-modal Decision Fusion for Continuous Authentication. *Computers and Electrical Engineering*, to appear, 2014.

[121] S. Acharya, A. Fridman, P. Brennan, P. Juola, R. Greenstadt, and M. Kam. User Authentication through Biometric Sensors and Decision Fusion. In *47th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2013.

[122] Hans Van Halteren, Harald Baayen, Fiona Tweedie, Marco Haverkort, and Anneke Neijt. New Machine Learning Methods Demonstrate the Existence of a Human Stylome. *Journal of Quantitative Linguistics*, 12(1):65–77, 2005.

[123] Ahmed Abbasi and Hsinchun Chen. Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace. *ACM Transactions on Information Systems (TOIS)*, 26(2):7, 2008.

[124] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.

[125] A Abbasi and H Chen. Identification and Comparison of Extremist-GroupWeb Forum Messages using Authorship Analysis. *IEEE Intelligent Systems*, 20(5):67–75, 2005.

[126] S. Acharya, A. Rajasekar, B. S. Shender, L. Hrebien, and M. Kam. Pulse Oximeter Signal Modeling and Fusion for Hypoxia Monitoring. In *17th International Conference on Information Fusion (FUSION)*, In Press.

[127] Arthur C Guyton. Textbook of Medical Physiology. *The American Journal of the Medical Sciences*, 242(2):136, 1961.

[128] Stanley R Crouch and James D Ingle. *Spectrochemical Analysis*. Prentice Hall, 1988.

[129] James E Sinex. Pulse Oximetry: Principles and Limitations. *The American Journal of Emergency Medicine*, 17(1):59–66, 1999.

[130] Joseph W Richardson. Aviation Hypoxia Monitor, December 13 1994. US Patent 5,372,134.

[131] Mark Kelly and Donald Pettit. Oxygen-Partial-Pressure Sensor for Aircraft Oxygen Mask. *NASA Tech Briefs*, 2003.

[132] Rajan Gurjar, Madhavi Seetamraju, David E Wolf, and John Hastings. High Reliability, Miniature Personal Hypoxia Monitoring System. In *SPIE Defense, Security, and Sensing*, pages 76740H–76740H, 2010. International Society for Optics and Photonics.

[133] B Shender, C Mattingly, M Warren, S Coleman, G Askew, and A Tucker. Relating the Time Complex Cognitive Performance Degrades to Physiologic Response During Moderate and Severe Normobaric Hypoxia. *Aviation, Space, and Environmental Medicine*, 84(4):340, 2013.

[134] Environics series 6202 reduced oxygen breathing device 2, Available: http://www.environics.com/robd.html.

[135] Steven M Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory (v. 1)*. Prentice Hall, 1993.

[136] K Sam Shanmugan and Arthur M Breipohl. *Random signals: Detection, Estimation, and Data Analysis*. Wiley, 1988.

[137] Brian DO Anderson and John B Moore. *Optimal Filtering*. Courier Dover Publications, 2012.

[138] Kedong Wang, Yong Li, and Chris Rizos. Practical Approaches to Kalman Filtering with Time-Correlated Measurement Errors. *IEEE Transactions on Aerospace and Electronic Systems,*, 48(2):1669–1681, 2012.

[139] V. Madisetti. *The Digital Signal Processing Handbook*. Electrical Engineering Handbook. Taylor & Francis, 1997.

[140] Robert F Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of united kingdom Inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

[141] Anil K Bera and Matthew L Higgins. Arch Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, 7(4):305–366, 1993.

Appendix A: Multi-Hypothesis Detection Probability Fusion

Krzystofowicz et al. in [26] proposed a Bayesian detection model for a cluster of distributed sensors where each sensor provides a detection probability instead of an observation vector or a local decision. The premise of the problem handled was fusion of such detection probabilities in a binary hypothesis space using Bayes' theorem. In this appendix a multiple hypothesis generalization of the same rule is derived.

Let $\Omega = \{H_1, ..., H_m\}$ be an exhaustive $m - ary$ hypothesis space on which decisions are to be made. A cluster of N sensors observe a situation and each sensor sends out a vector of detection probabilities for each of the hypothesis. For the i^{th} sensor, the vector would be represented as

$$G_i = [g_i(H_1), ..., g_i(H_m)]^T$$

with $i = 1, ...N$. For convenience sake, we drop the argument H_i and represent G_i as $[g_{1i}, ..., g_{mi}]^T$. Note that, in every G_i, each element is a detection probability on an exhaustive set and hence the following holds $\forall i = 1, ..., N$

$$\sum_{j=1}^{m} g_{ji} = 1. \tag{A.1}$$

Every vector G_i contains m components which are non-negative and satisfy (A.1). This makes the set of all such vectors form a $m-1$ dimensional simplex residing in $\mathbb{R}^m$.

Each point G_i in the simplex can be thought of as a probability mass function since its components are non-negative and sum to 1. The Dirichlet distribution defined as a distribution over probability mass functions of length m can be used to model the vectors G_i, $i = 1, ..., N$. For each vector G_i, let $\alpha_i = [\alpha_{1i}, ..., \alpha_{mi}]$ be the parameter vector with $\alpha_{0i} = \sum_{j=1}^{m} \alpha_{ji}$. Then G_i is Dirichlet distributed with the parameter set α_i and the distribution function given as

$$f_i(G_i; \alpha_i) = \frac{\Gamma(\alpha_{0i})}{\prod_{j=1}^{m} \Gamma(\alpha_{ji})} \prod_{j=1}^{m} (g_{ji})^{\alpha_{ji}-1}. \tag{A.2}$$

Dirichlet distribution is the multivariate generalization of the Beta distribution and therefore the marginal distributions of the components of the vector G_i will be Beta distributed as follows

$$g_{ij} \sim Beta(\alpha_{ji}, \alpha_{0i} - \alpha_{ji}). \tag{A.3}$$

Let us define the $m \times N$ matrix G as

$$G = [G_1 \;\; G_2 \;\; ... \;\; G_N].$$

Expanding the matrix G, we have

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1N} \\ g_{21} & g_{22} & \cdots & g_{2N} \\ & \vdots & & \\ g_{m1} & g_{m2} & \cdots & g_{mN} \end{bmatrix} \tag{A.4}$$

We note that each row of the matrix G represents the detection probabilities obtained from all the N sensors toward a particular hypothesis which can be identified by the row number. Let us assume the prior probabilities of the m hypotheses arranged in order to be given by $P(H_1), P(H_2), ..., P(H_m)$. For an arbitrary row, say j^{th}, $j = 1, ..., m$ row of the matrix G, the joint probability distribution would be given by

$$\xi_j(g_{j1}, g_{j2}, ..., g_{jN}) = \sum_{k=1}^{m} f(g_{j1}, g_{j2}, ..., g_{jN}|H_k)P(H_k). \tag{A.5}$$

Using Bayes' rule, the posterior detection probability of hypothesis H_j conditioned on the detection probabilities $g_{1j}, g_{2j}, ..., g_{nj}$ can be obtained as follows

$$\begin{aligned} \eta_j(g_{j1}, g_{j2}, ..., g_{jN}) &= P(H_j|g_{j1}, g_{j2}, ..., g_{jN}) \\ &= \frac{f(g_{j1}, g_{2j}, ..., g_{jN}|H_j)P(H_j)}{\sum_{k=1}^{m} f(g_{j1}, g_{j2}, ..., g_{jN}|H_k)P(H_k)} \\ &= \left[1 + \frac{\sum_{k\neq j}^{m} f(g_{k1}, g_{k2}, ..., g_{kN}|H_k)P(H_k)}{f(g_{j1}, g_{j2}, ..., g_{jN}|H_j)P(H_j)}\right]^{-1} \end{aligned} \tag{A.6}$$

Assuming the local detection probabilities are independent conditioned on the hypotheses, we have

$$f(g_{j1}, g_{j2}, ..., g_{jN}|H) = \prod_{i=1}^{N} f(g_{ji}|H). \tag{A.7}$$

Using (A.7) in (A.6) we have

$$\eta_j(g_{j1}, g_{j2}, ..., g_{jN}) = \left[1 + \frac{\sum_{k\neq j}^{m} \left[\prod_{i=1}^{N} f(g_{ki}|H_k)P(H_k)\right]}{\prod_{i=1}^{N} f(g_{ji}|H_j)P(H_j)}\right]^{-1}. \tag{A.8}$$

The net resultant posterior detection probability vector would then be

$$\eta = [\eta_1, \eta_2, ..., \eta_m]^T. \tag{A.9}$$

Vita

Sayandeep Acharya received his undergraduate education (B. Tech, Electronics and Communication Engineering, 2006) from West Bengal University of Technology, India. From 2006 to 2007 he was associated with Wipro Technologies, India as a software engineer. He received M.S., in Electrical Engineering from Drexel University in 2009. He joined the Data Fusion Laboratory at Drexel University in 2009 under the supervision of Dr. Moshe Kam working toward the Ph.D., degree. His research involves investigation and development of multi-sensor fusion methodologies for distributed detection systems. His other research interests include signal modeling, applied machine learning, classification and pattern recognition.