# Fusion in the Context of Information Theory

Mohiuddin Ahmed and Gregory Pottie
mohin@ee.ucla.edu, pottie@ee.ucla.edu
Electrical Engineering Department
University of California, Los Angeles

# Chapter 3

# Information Fusion

## 3.9 Fusion in the Context of Information Theory

In this section, we selectively explore some aspects of the theoretical framework that has been developed to analyze the nature, performance, and fundamental limits for information processing in the context of data fusion. In particular, we discuss how Bayesian methods for distributed data fusion can be interpreted from the point of view information theory. Consequently, information theory can provide a common framework for distributed detection and communication tasks in sensor networks.

Initially, the context is established for considering distributed networks as efficient information processing entities (Section 3.9.1). Next, in Section 3.9.2, the approaches taken towards analyzing such systems and the path leading towards the modern information theoretic framework for information processing are discussed. The details of the mathematical method are highlighted in Section 3.9.3, and applied specifically for the case of multi-sensor systems in Section 3.9.4. Finally our conclusions are presented in Section 3.9.5.

### 3.9.1 Information Processing in Distributed Networks

Distributed networks of sensors and communication devices provide the ability to electronically network together what were previously isolated islands of information sources and sinks, or more generally, *states of nature*. The states can be measurements of physical parameters (e.g. temperature, humidity, etc.) or estimates of operational conditions (network loads, throughput, etc.), among other things, distributed over a region in time and/or space. Previously, the aggregation, fusion and interpretation of this mass of data representing some phenomena of interest were performed by isolated *sensors*, requiring human supervision and control. However, with the advent of powerful hardware platforms and networking technologies, the possibility and advantages of *distributed sensing* information processing has been recognized [32].

A sensor can be defined to be any device that provides a quantifiable set of outputs in response to a specific set of inputs. These outputs are useful if they can be mapped to a state of nature that is under consideration. The end goal of the sensing task is to acquire a description of the external world, predicated upon which can be a series of *actions*. In this context, sensors can be thought of as *information gathering, processing and dissemination* entities, as diagrammed in Figure 3.9.1. The data pathways in the figure illustrate an

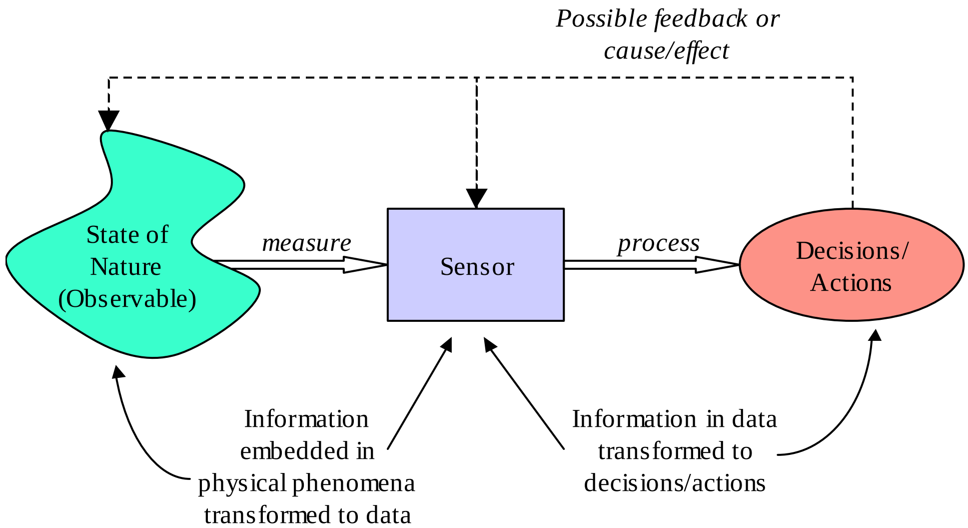


Figure 3.9.1: Information Processing in Sensors

abstraction of the flow of information in the system. In a distributed network of sensors, the sensing system may be comprised of multiple sensors that are physically disjoint or distributed in time or space, and that work cooperatively. Compared to a single sensor platform, a network has the advantages of *diversity* (different sensors offer complementary viewpoints), and *redundancy* (reliability and increased resolution of the measured quantity) [24]. In fact, it has been rigorously established from the theory of distributed detection that higher reliability and lower probability of detection error can be achieved when observation data from multiple, distributed sources is intelligently fused in a decision making algorithm, rather than using a single observation data set [44]. Intuitively, any practical sensing device has limitations on its sensing capabilities (e.g. resolution, bandwidth, efficiency, etc.). Thus, descriptions built on the data sensed by a single device are only approximations of the true state of nature. Such approximations are often made worse by incomplete knowledge and understanding of the environment that is being sensed and its interaction with the sensor. These uncertainties, coupled with the practical reality of occasional sensor failure greatly compromises reliability and reduces confidence in sensor measurements. Also, the spatial and physical limitations of sensor devices often means that only partial information can be provided by a single sensor.

A network of sensors overcomes many of the shortcomings of a single sensor. However new problems in efficient information management arise. These may be categorized into two broad areas [30]:

1. *Data Fusion:* This is the problem of combining diverse and sometimes conflicting information provided by sensors in a multi-sensor system, in a consistent and coherent manner. The objective is to infer the relevant states of the system that is being observed or activity being performed.

2. *Resource Administration:* This relates to the task of optimally configuring, coordinating and utilizing the available sensor resources, often in a dynamic, adaptive environment. The objective is to ensure efficient[1] use of the sensor platform for the task at hand.

In comparison to lumped-parameter sensor systems (Figure 3.9.1), the issues mentioned

[1] *Efficiency*, in this context, is very general and can refer to power, bandwidth, overhead, throughput, or a variety of other performance metrics, depending upon the particular application.

above for multi-sensor systems can be diagrammed as shown in Figure 3.9.2 [24].

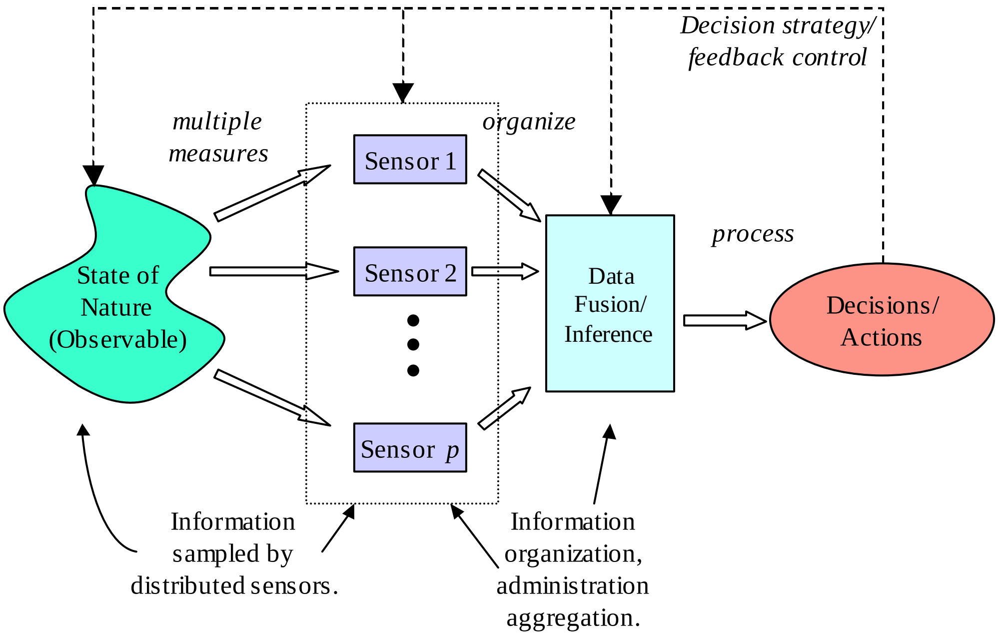


Figure 3.9.2: Information Processing in Distributed Sensors

### 3.9.2 Evolution Towards Information Theoretic Methods for Data Fusion

Most of the early research effort in probabilistic and information theoretic methods for data fusion focused on techniques motivated by specific applications, such as in vision systems, sonar, robotics platforms, etc. [19, 23, 26, 22]. As the inherent advantages of using multi-sensor systems were recognized [46, 2], a need for a comprehensive theory of the associated problems of distributed, decentralized data fusion and multi-user information theory became apparent [43, 15, 7]. Advances in integrated circuit technology have enabled mass production of sensors, signal processing elements and radios [12, 34], spurring new research in wireless communications [20], and in *ad hoc networking* [36, 40]. Subsequently, it was only natural to combine these two disciplines—sensors and networking—to develop a new generation of distributed sensing devices that can work cooperatively to exploit diversity [31, 32]. An abridged overview of the research in sensor fusion and management is now summarized [24].

*Sensor Fusion Research:* Data fusion is the process by which data from a multitude of sensors is used to yield an optimal estimate of a specified state vector pertaining to the observed system [44], whereas sensor administration is the design of communication and control mechanisms for the efficient use of distributed sensors, with regards to power, performance, reliability, etc. Data fusion and sensor administration have mostly been addressed separately. Sensor administration has been addressed in the context of wireless networking and not necessarily in conjunction with the unique constraints imposed by data fusion methodologies.

To begin with, sensor models have been aimed at interpretation of measurements. This approach to modeling can be seen in the sensor models used by Kuc and Siegel [19], among others. Probability theory, and in particular, a Bayesian treatment of data fusion [9] is arguably the most widely used method for describing uncertainty in a way that abstracts from a sensor's physical and operational details. Qualitative methods have also been used to describe sensors, for example by Flynn [11] for sonar and infra-red applications. Much work has also been done in developing methods for intelligently combining information from different sensors. The basic approach has been to pool the information using what are essentially "weighted averaging" techniques of varying degrees of complexity. For example Berger [2] discusses a majority voting technique based on a probabilistic representation of information. Non-probabilistic methods [16] used inferential techniques, for example for multi-sensor target identification. Inferring the state of nature given a probabilistic representation is, in general, a well understood problem in classical estimation. Representative methods are Bayesian estimation, Least Squares estimation, Kalman Filtering, and its various derivatives. However, the question of how to use these techniques in a distributed fashion has not been addressed to date in a systematic fashion except for some specific physical layer cases [45].

*Sensor Administration Research:*

In the area of sensor network administration, protocol development and management have mostly been addressed using application specific descriptive techniques for specialized systems [46]. Tracking radar systems provided the impetus for much of the early work. Later, robotic applications led to the development of models for sensor behavior and performance that could then be used to analyze and manage the transfer of sensor data. The centralized or hierarchical nature of such systems enabled this approach to succeed. Other schemes that found widespread use were based on determining cost functions and performance trade-offs a priori [1], e.g. cost-benefit assignment matrices allocating sensors to targets, or Boolean matrices characterizing sensor-target assignments based on sensor availability and capacity. Expert system approaches have also been used, as well as decision-theoretic (*normative*) techniques. However, optimal sensor administration in this way has been shown by Tsitsiklis [43] to be very hard in the general framework of distributed sensors, and practical schemes use a mixture of heuristic techniques (for example in data fusion systems involving wired sensors in combat aircrafts). Only recently have the general networking issues for wireless ad hoc networks been addressed (Sohrabi, Singh [41, 39]), where the main problems of self-organization, bootstrap, route discovery etc., have been identified. Application specific studies, e.g. in the context of antenna arrays (Yao, [47]) have also discussed these issued. However, few general fusion rules or data aggregation models for networked sensors have been proposed, with little analytical or quantitative emphasis. Most of these studies do not analyze in detail the issues regarding the network-global impact of administration decisions, such as choice of fusion nodes, path/tree selections, data fusion methodology, or physical layer signalling details.

### 3.9.3 Probabilistic Framework for Distributed Processing

The information being handled in multi-sensor systems almost always relates to a state of nature, and consequently, it is assumed to be unknown prior to observation or estimation. Thus, the model of the information flow shown in Figure 3.9.2 is probabilistic, and hence can be quantified using the principles of information theory [6, 14]. Furthermore, the process of data detection and processing that occurs within the sensors and fusion node(s) can be considered as elements of classical statistical decision theory [29]. Using the mature techniques that these disciplines offer, a probabilistic information processing relation can be quantified for sensor networks, and analyzed within the framework of the well-known Bayesian paradigm [35]. The basic tasks in this approach are the following:

1. Determination of appropriate information processing techniques, models and metrics for fusion and sensor administration.

2. Representation of the sensors process, data fusion, and administration methodologies using the appropriate probabilistic models.

3. Analysis of the measurable aspects of the information flow in the sensor architecture using the defined models and metrics.

4. Design of optimal data fusion algorithms and architectures for optimal inference in multi-sensor systems.

5. Design, implementation and test of associated networking and physical layer algorithms and architectures for the models determined in (4).

We now consider two issues in information combining in multi-sensor systems: **(i)** the nature of the information being generated by the sensors, and **(ii)** the method of combining the information from disparate sources.

### Sensor Data Model for Single Sensors

Any observation or measurement by any sensor is always uncertain to a degree determined by the *precision* of the sensor. This uncertainty, or measurement *noise*, requires us to treat the data generated by a sensor probabilistically. We therefore adopt the notation and definitions of probability theory to determine an appropriate model for sensor data [13, 24].

**Definiton 3.9.1.** A *state vector* at time instant $t$, is a representation of the *state of nature* of a process of interest, and can be expressed as a vector $\boldsymbol{x}(t)$ in a measurable, finite-dimensional vector space, $\Omega$, over a discrete or continuous field, $\mathcal{F}$:

$$\boldsymbol{x}(t) \in \Omega \subseteq \mathbb{R}^n \tag{3.9.1}$$

The state vector is arbitrarily assumed to $n$-dimensional and can represent a particular state of nature of interest, e.g. it can be the three dimensional position vectors of an airplane. The state space may be either continuous or discrete (e.g. the on or off states of a switch).

**Definiton 3.9.2.** A *measurement vector* at time instant $t$ is the information generated by a single sensor (in response to an observation of nature), and can be represented by an $m$-dimensional vector, $\boldsymbol{z}(t)$ from a measurement vector space $\Psi$.

$$\boldsymbol{z}(t) = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \in \Psi \subseteq \mathbb{R}^m \tag{3.9.2}$$

Intuitively, the measurement vector may be thought of as $m$ pieces of data that a single sensor generates from a single observation at a single instant of time. Because of measurement error, the sensor output $\boldsymbol{z}(t)$ is an approximation of $\boldsymbol{x}(t)$—the true state of nature. It is important to note that $\boldsymbol{z}(t)$ may itself not be directly visible to the user of the sensor platform. A noise corrupted version $\Gamma\{\boldsymbol{z}(t), \boldsymbol{v}(t)\}$, as defined below, may be all that is available for processing. Furthermore, the dimensionality of the sensor data may not be the same as the dimension of the observed parameter that is being measured. For example, continuing with the airplane example, a sensor may display the longitude and latitude of the airplane at a particular instant of time via GPS (a 2-dimensional observation vector), but may not be able to measure the altitude of the airplane (which completes the 3-dimensional specification of the actual location of the airplane in space).

The *measurement error* itself can be considered as another vector, $\boldsymbol{v}(t)$, or a *noise process* vector, of the same dimensionality as the observation vector $\boldsymbol{z}(t)$. As the name suggests, noise vectors are inherently stochastic in nature, and serve to render all sensor measurements uncertain, to a specific degree.

**Definiton 3.9.3.** An *observation model*, $\Gamma$, for a sensor is a mapping from state space $\Omega$ to observation space $\Psi$, and is parameterized by the statistics of the noise process:

$$\Gamma_{\boldsymbol{v}} : \Omega \mapsto \Psi. \tag{3.9.3}$$

Functionally, the relationship between the state, observation and noise vectors can be expressed as:

$$\boldsymbol{z}(t) = \Gamma\left\{\boldsymbol{x}(t), \boldsymbol{v}(t)\right\}. \tag{3.9.4}$$

**Objective:** The objective in sensing applications is to infer the unknown state vector $\boldsymbol{x}(t)$ from the error corrupted and (possibly lower dimensional) observation vector $\boldsymbol{z}(t), \boldsymbol{v}(t)$. If the functional specification of the mapping in Equation (3.9.3), and the noise vector $\boldsymbol{v}(t)$, were known for all times $t$, then finding the inverse mapping for one-to-one cases would be trivial, and the objective would be easily achieved. It is precisely because either or both parameters may be random that gives rise to various estimation architectures for inferring the state vector from the imperfect observations. A geometric interpretation of the objective can be presented as shown in Figure 3.9.3(i). The simplest mapping relationship $\Gamma$ that can be used as a sensor data model is the *additive* model of noise corruption, as shown in Figure 3.9.3(ii), which can be expressed as:

$$\boldsymbol{x} = \Gamma\left(\boldsymbol{z} + \boldsymbol{v}\right). \tag{3.9.5}$$

Typically, for well designed and matched sensor platforms, the noise vector is small compared

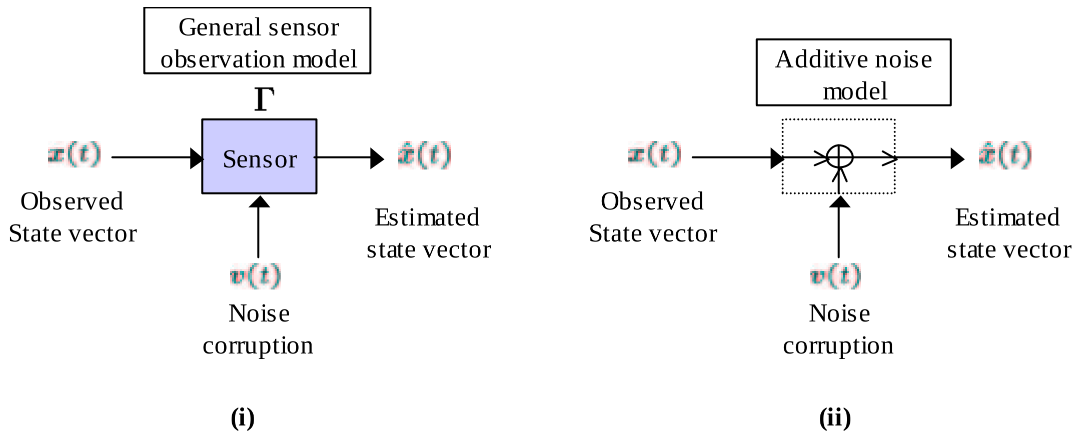


Figure 3.9.3: Sensor data models: (i) General case (ii) Noise additive case.

to the measurement vector, in which case a Taylor approximation can be made:

$$\boldsymbol{x} = \Gamma(\boldsymbol{z}) + (\nabla_{\boldsymbol{z}}\Gamma)\, \boldsymbol{z} + (\text{higher order terms}) \tag{3.9.6}$$

where $\nabla_{\boldsymbol{z}}$ is the Jacobian matrix of the mapping $\Gamma$ with respect to the state measurement vector $\boldsymbol{z}$. Since the measurement error is random, the state vector observed is also random, and we are in essence dealing with random variables. Thus, we can use well established statistical methods to quantify the uncertainty in the random variables [35]. For example, the statistics of the noise process $\boldsymbol{v}(t)$ can be often be known *a priori*. Moments are the most commonly used measures for this purpose, and in particular, if the covariance of the noise process is known, $\mathbf{E}\left\{\boldsymbol{v}\boldsymbol{v}^T\right\}$, then the covariance of the state vector is [24]:

$$\mathbf{E}\left\{\boldsymbol{x}\boldsymbol{x}^T\right\} = (\nabla_{\boldsymbol{z}}\Gamma)\, \mathbf{E}\left\{\boldsymbol{v}\boldsymbol{v}^T\right\} (\nabla_{\boldsymbol{z}}\Gamma)^T . \tag{3.9.7}$$

For uncorrelated noise $\boldsymbol{v}$, the matrix $(\nabla_{\boldsymbol{z}}\Gamma)\, \mathbf{E}\left\{\boldsymbol{v}\boldsymbol{v}^T\right\} (\nabla_{\boldsymbol{z}}\Gamma)^T$ is symmetric and can be decomposed using singular value decomposition [37]:

$$(\nabla_{\boldsymbol{z}}\Gamma)\, \mathrm{E}\left\{\boldsymbol{v}\boldsymbol{v}^T\right\} (\nabla_{\boldsymbol{z}}\Gamma)^T = \left(\mathsf{SDS}^T\right) \tag{3.9.8}$$

where $\mathsf{S}$ is an $(n \times n)$ matrix of orthogonal vectors $\boldsymbol{e}_j$ and D are the eigenvalues of the decomposition:

$$\mathsf{S} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n), \quad \boldsymbol{e}_i\boldsymbol{e}_j = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases} \tag{3.9.9}$$

$$\mathsf{D} = diag\,(d_1, d_2, \ldots, d_n) \tag{3.9.10}$$

The components of $\mathsf{D}$ correspond to the scalar variance in each of direction. Geometrically, all the directions for a given state $\boldsymbol{x}$ can be visualized as an ellipsoid in $n$-dimensional space, with the principal axes in the directions of the vectors $\boldsymbol{e}_k$ and $2\sqrt{d_j}$ as the corresponding magnitudes. The volume of the ellipsoid is the uncertainty in $\boldsymbol{x}$. The 2-dimensional case is shown in Figure 3.9.4. From this perspective, the basic objective in the data fusion

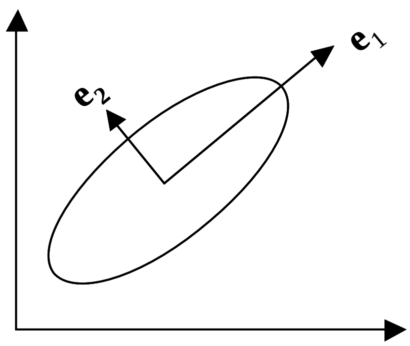


Figure 3.9.4: Ellipsoid of state vector uncertainty

problem is then to reduce the volume of the uncertainty ellipsoid. All the techniques for data estimation, fusion, and inference are designed towards this goal [27].

### A Bayesian Scheme for Decentralized Data Fusion

Given the inherent uncertainty in measurements of states of nature, the end goal in using sensors, as mentioned in the previous section, is to obtain the best possible estimates of the states of interest for a particular application. The Bayesian approach to solving this problem is concerned with quantifying *likelihoods* of events, given various types of partial knowledge or observations, and subsequently determining the state of nature that is most probably responsible for the observations as the 'best' estimate.

The issue of whether the Bayesian approach is intrinsically the 'best' approach for a particular problem[2] is a philosophical debate that is not discussed here further. It may be mentioned, however, that arguably, the Bayesian paradigm is most *objective* because it is based only on observations and 'impartial' models for sensors and systems.

The information contained in the (noise corrupted) measured state vector $\boldsymbol{z}$ is first described by means of probability distribution functions (PDF). Since all observations of states of nature are causal manifestations of the underlying processes governing the state of nature[3], the PDF of $\boldsymbol{z}$ is conditioned by the state of nature at which time the observation/measurement was made. Thus, the PDF of $\boldsymbol{z}$ conditioned by $\boldsymbol{x}$ is what is usually measurable and is represented by:

$$F_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{x}) \tag{3.9.11}$$

This is known as the *Likelihood Function* for the observation vector. Next, if information about the possible states under observation is available (e.g. *a priori* knowledge of the range of possible states), or more precisely the probability distribution of the possible states $F_{\boldsymbol{X}}(\boldsymbol{x})$, then the prior information and the likelihood function (3.9.11) can be combined to provide the *a posteriori* conditional distribution of $\boldsymbol{x}$, given $\boldsymbol{z}$, by Bayes' Theorem [13]:

**Theorem 3.9.1.**

$$F_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{z}) = \frac{F_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x})}{\displaystyle\int_{\Omega} F_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}F(\boldsymbol{x})} = \frac{F_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x})}{F_{\boldsymbol{Z}}(\boldsymbol{z})} \tag{3.9.12}$$

Usually, some function of the actual likelihood function, $g(T(\boldsymbol{z}) \mid \boldsymbol{x})$, is commonly available as the processable information from sensors. $T(\boldsymbol{z})$ is known as the *sufficient statistic*

[2]In contrast with various other types of inferential and subjective approaches [35]

[3]Ignoring the observer-state interaction difficulties posed by Heisenberg Uncertainty considerations.

for $\boldsymbol{x}$ and Equation (3.9.12) can be reformulated as:

$$F_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{z}) = F_{\boldsymbol{X}}(\boldsymbol{x} \mid T(\boldsymbol{z})) = \frac{g(T(\boldsymbol{z}) \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x})}{\displaystyle\int_{\Omega} g(T(\boldsymbol{z}) \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}F(\boldsymbol{x})} \tag{3.9.13}$$

When observations are carried out in discrete time steps according to a desired resolution, then a vector formulation is possible. Borrowing notation from [24], all observations up to time index $r$ can be defined as:

$$\boldsymbol{Z}^r \triangleq \{\boldsymbol{z}(1), \boldsymbol{z}(2), \ldots, \boldsymbol{z}(r)\} \tag{3.9.14}$$

from where the posterior distribution of $\boldsymbol{x}$ given the set of observations $\boldsymbol{Z}^r$ becomes:

$$F_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{Z}^r) = \frac{F_{\boldsymbol{Z}^r}(\boldsymbol{Z}^r \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x})}{F_{\boldsymbol{Z}^r}(\boldsymbol{Z}^r)}. \tag{3.9.15}$$

Using the same approach, a recursive version of Equation (3.9.15) can also be formulated:

$$F_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{Z}^r) = \frac{F_{\boldsymbol{Z}}(\boldsymbol{z}(r) \mid \boldsymbol{x}) F_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{Z}^{r-1})}{F_{\boldsymbol{Z}}(\boldsymbol{z}(r) \mid \boldsymbol{Z}^{r-1})} \tag{3.9.16}$$

in which case all the $r$ observations do not need to be stored, and instead only the current observation $\boldsymbol{z}(r)$ can be considered at the $r^{th}$ step. This version of the Bayes' Law is most prevalent in practice since it offers a directly implementable technique for fusing observed information with *prior beliefs*.

*Classical Estimation Techniques*

A variety of inference techniques can now be applied to estimate the state vector $\boldsymbol{x}$ (from the time series observations from a single sensor). The estimate, denoted by $\hat{\boldsymbol{x}}$, is derived from the posterior distribution $F_{vecx}(\boldsymbol{x} \mid \boldsymbol{Z}^r)$ and is a point in the uncertainty ellipsoid of Figure 3.9.4. The basic objective is to reduce the volume of the ellipsoid, which is equivalent to minimizing the probability of error based on some criterion. Three classical techniques are now briefly reviewed: *Maximum Likelihood, Maximum A Posteriori* and *Minimum Mean Square Error* estimation.

*Maximum Likelihood* (ML) estimation involves maximizing the likelihood function (Equation 3.9.11) by some form of search over the state space $\Omega$:

$$\hat{\boldsymbol{x}}_{ML} = \arg\max_{\boldsymbol{x}\in\Omega} F_{\boldsymbol{Z}^r}\left(\boldsymbol{Z}^r \mid \boldsymbol{x}\right) \tag{3.9.17}$$

This is intuitive since the PDF is greatest when the correct state has been guessed for the conditioning variable. However, a major drawback is that for state vectors from large state spaces, the search may be computationally expensive, or infeasible. Nonetheless, this method is widely used in many disciplines, e.g. digital communication reception [33].

*Maximum a posteriori* (MAP) estimation technique involves maximizing the posterior distribution from observed data as well as from prior knowledge of the state space:

$$\hat{\boldsymbol{x}}_{MAP} = \arg\max_{\boldsymbol{x}\in\Omega} F_{\boldsymbol{x}}\left(\boldsymbol{x} \mid \boldsymbol{Z}^r\right) \tag{3.9.18}$$

Since prior information may be subjective, objectivity for an estimate (or the inferred state) is maintained by considering only the likelihood function (i.e. only the observed information). In the instance of no prior knowledge, and the state space vectors are all considered to be equally likely, the MAP and ML criterion can be shown to be identical.

*Minimum Mean Square Error* (MMSE) techniques attempt to minimize the estimation error by searching over the state space, albeit in an organized fashion. This is the most popular technique in a wide variety of information processing applications, since the variable can often be found analytically, or the search space can be reduced considerably or investigated systematically. The key notion is to reduce the covariance of the estimate. Defining the mean and variance of the posterior observation variable as:

$$\bar{\boldsymbol{x}} \triangleq \mathrm{E}_{F(\boldsymbol{x}|\boldsymbol{Z}^r)}\{x\} \tag{3.9.19}$$

$$\mathrm{Var}(\boldsymbol{x}) \triangleq \mathrm{E}_{F(\boldsymbol{x}|\boldsymbol{Z}^r)}\{(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\} \tag{3.9.20}$$

it can be shown that the least squares estimator is one that minimizes the Euclidean distance between the true state $\boldsymbol{x}$ and the estimate $\hat{\boldsymbol{x}}$, given the set of observations $\boldsymbol{Z}^r$. In the context of random variables, this estimator is referred to as the MMSE estimate and can be expressed as:

$$\hat{\boldsymbol{x}}_{MMSE} = \arg\min_{\boldsymbol{x}\in\Omega} \mathrm{E}_{F(\boldsymbol{x}|\boldsymbol{Z}^r)}\{(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\} \tag{3.9.21}$$

To obtain the minimizing estimate, Equation (3.9.21) can be differentiated with respect to $\hat{\boldsymbol{x}}$ and set equal to zero, which yields $\hat{\boldsymbol{x}} = \mathrm{E}\{\boldsymbol{x} \mid \boldsymbol{Z}^r\}$. Thus the MMSE estimate is the conditional mean. It also can be shown that the MMSE estimate is the minimum variance estimate, and when the conditional density coincides with the mode, the MAP and MMSE estimators are equivalent.

These estimation techniques and their derivatives such as the Wiener and Kalman filters [18] all serve to reduce the uncertainty ellipsoid associated with state $\boldsymbol{x}$ [27]. In fact, direct applications of these mathematical principles formed the field of radio frequency *signal detection* in noise, and shaped the course of developments in digital communication technologies.

### Distributed Detection Theory and Information theory

Information theory was developed to determine the fundamental limits on the performance of communication systems [38]. Detection theory on the other hand, involves the application of statistical decision theory to estimate states of nature, as discussed in the previous section. Both these disciplines can be used to treat problems in the transmission and reception of information, as well as the more general problem of data fusion in distributed systems. The synergy was first explored by researchers in the 1950s and 1960s [25], and the well established source and channel coding theories were spawned as a result. With respect to data fusion, the early research in the fields of information theory and fusion proceeded somewhat independently. Whereas information theory continued exploring the limits of digital signalling, data fusion, on the other hand, and its myriad ad hoc techniques were developed by the practical concerns of signal detection, aggregation and interpretation for decision making. Gradually, however, it was recognized that both issues, at their abstract levels, dealt fundamentally with problems of information processing.

Subsequently, attempts were made to unify distributed detection and fusion theory, as it applied e.g. in sensor fusion, with the broader field of information theory. Some pioneering work involved the analysis of the hypothesis testing problem using discrimination (Kullback, [21]), employing cost functions based on information theory for optimizing signal detection (Middleton, [25]) and formulating the detection problem as a coding problem for asymptotic analysis using error exponent functions (Csiszar et al. [8], Blahut[3]). More recently, research in these areas have been voluminous, with various theoretical studies exploring the performance limits and asymptotic analysis of fusion and detection schemes ([43, 4]).

In particular, some recent results [44] are relevant to the case of a distributed system of sensor nodes. As has been noted earlier, the optimal engineering trade-offs for the efficient design for such a system is not always clear cut. However, if the detection/fusion problem can be recast in terms of information theoretic cost functions, then it has been shown that system optimization techniques provide useful design paradigms.

For example consider the block diagrams of a conventional binary detection system and a binary communication channel as shown in Figure 3.9.5. The source in the detection problem can be viewed as the information source in the information transmission problem. The decisions in the detection model can be mapped as the channel outputs in the channel model. Borrowing the notation from [44], if the input is considered a random variable $H = i,\ i = 0, 1$ where probability $P(H = 0) = P_0$, the output $u = i, i = 0, 1$ is then a decision random variable, whose probabilities of detection ($P_D$), miss ($P_M$), false alarm

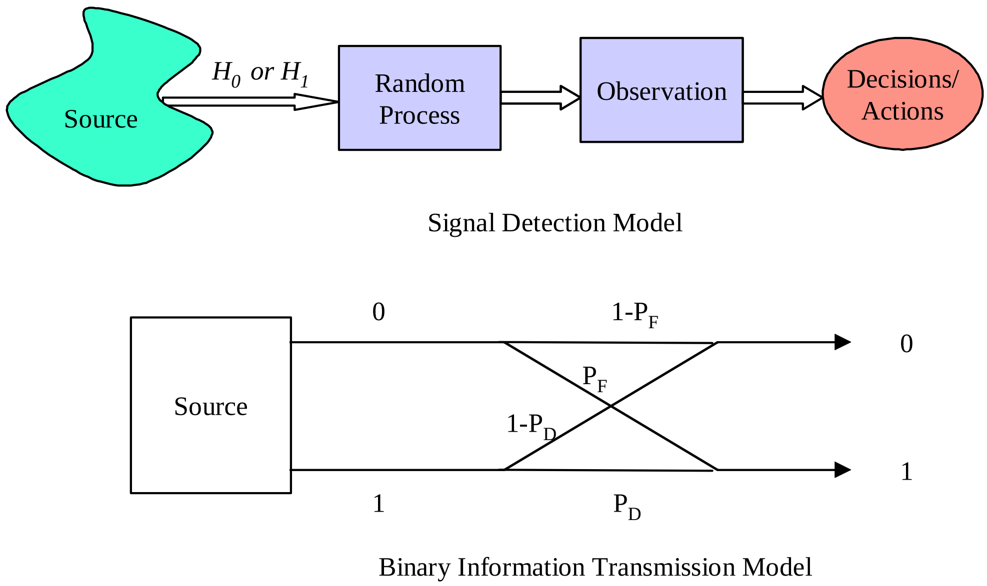


Figure 3.9.5: Signal Detection vs. Information Transmission

($P_F$), etc. can be interpreted in terms of the transition probabilities of the information transmission problem. This is the classic example of the binary channel [33].

If the objective of the decision problem is the minimization of the information loss between the input and output, then it can be shown that the objective is equivalent to the maximization of the *mutual information*, $I(H;u)$ (see Section 3.9.4 for formal definitions of entropy and information measures). This provides a mechanism for computing practical likelihood test ratios as a technique for *information-optimal* data fusion. Thus, the case of the binary detection problem, the *a posteriori* probabilities are:

$$P(u=0) = P_0(1-P_F) + (1-P_0)(1+P_D) \qquad \triangleq \alpha_0 \tag{3.9.22}$$

$$P(u=1) = P_0 P_F + (1-P_0)P_D \qquad \triangleq \alpha_1 \tag{3.9.23}$$

whereupon it can be shown that the optimal decision threshold for the received signal is:

$$threshold = \frac{-P_0\left\{\log(\alpha_0/\alpha_1) - \log\left[(1-P_F)/P_F\right]\right\}}{(1-P_0)\left\{\log(\alpha_0/\alpha_1) - \log\left[(1-P_D)/P_D\right]\right\}} \tag{3.9.24}$$

This approach can be extended to the case of distributed detection. For example, for a detection system in a parallel topology without a fusion center, and assuming the observations at the local detectors are conditionally independent, the goal is then to maximize the mutual information $I(H;\boldsymbol{u})$ where the vector $\boldsymbol{u}$ contains the local decisions. Once again, it can be shown that the optimal detectors are threshold detectors, and likelihood ratio tests can then be derived for each detector. Using the second subscript in the variables below to refer to the detector number, the thresholds are:

$$threshold_1 = -\frac{P_0\left[\log\left(\frac{\alpha_{00}}{\alpha_1 0}\right) + P_{F2}\log\left(\frac{\alpha_{01}\alpha 10}{\alpha_{00}\alpha 11}\right) - \log\left(\frac{1-P_{F1}}{P_{F1}}\right)\right]}{(1-P_0)\left[\log\left(\frac{\alpha_{00}}{\alpha_1 0}\right) + P_{D2}\log\left(\frac{\alpha_{01}\alpha 10}{\alpha_{00}\alpha 11}\right) - \log\left(\frac{1-P_{D1}}{P_{D1}}\right)\right]} \tag{3.9.25}$$

with a similar expression for $threshold_2$. In a similar manner, other entropy-based information theoretic criteria (e.g. logarithmic cost functions) can be successfully used to design the detection and distributed fusion rules in an integrated manner for various types of fusion architectures, (e.g. serial, parallel with fusion center, etc.) This methodology provides an attractive, unified approach for system design, and has the intuitive appeal of treating the distributed detection problem as an information transmission problem.

### 3.9.4 Bayesian Framework for Distributed Multi-Sensor Systems

When a number of spatially and functionally different sensor systems are used to observe the same (or similar) state of nature, then the data fusion problem is no longer simply a state space uncertainty minimization issue. The distributed and multi-dimensional nature of the problem requires a technique for checking the usefulness and validity of the data from each of the not necessarily independent sensors. The data fusion problem is more complex, and general solutions are not readily evident. This section explores some of the commonly studied techniques and proposes a novel, simplified methodology that achieves some measure of generality.

The first issue is the proper modeling of the data sources. If there are $p$ sensors observing the same state vector, but from different vantage points, and each one generates its own observations, then we have a collection of observation vectors $\boldsymbol{z}_1(t)$, $\boldsymbol{z}_2(t), \ldots, \boldsymbol{z}_p(t)$, which can be represented as a combined matrix of all the observations from all sensors (at any particular time $t$):

$$\boldsymbol{Z}(t) = \begin{pmatrix} \boldsymbol{z}_1(t) & \boldsymbol{z}_2(t) & \cdots & \boldsymbol{z}_p(t) \end{pmatrix} = \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{p1} \\ z_{12} & z_{22} & \cdots & z_{p2} \\ & & \ddots & \\ z_{1m} & z_{2m} & \cdots & z_{pm} \end{bmatrix}. \tag{3.9.26}$$

Furthermore, if each sensor makes observations up to time step $r$ for a discretized (sampled) observation scheme, then the matrix of observations $\boldsymbol{Z(r)}$ can be used to represent the observations of all the $p$ sensors at time-step $r$ (a discrete variable, rather than the continuous $\boldsymbol{Z}(t)$). With adequate memory allocation for signal processing of the data, we can consider the super-matrix $\{\boldsymbol{Z}^r\}$ of all the observations of all the $p$ sensors from time step 0 to $r$:

$$\{\boldsymbol{Z}^r\} = \bigcup_{i=1}^{p} \boldsymbol{Z}_i^r \tag{3.9.27}$$

$$\text{where} \quad \boldsymbol{Z}_i^r = \{\boldsymbol{z}_i(1), \boldsymbol{z}_i(2), \ldots \boldsymbol{z}_i(r)\} \tag{3.9.28}$$

This suggests that to use all the available information for effectively fusing the data from multiple sensors, what is required is the global posterior distribution $F_{\boldsymbol{x}}(\boldsymbol{x} \mid \{\boldsymbol{Z}^r\})$, given the time-series information from each source. This can be accomplished in a variety of ways, the most common of which are summarized below [24].

The **Linear Opinion Pool** [42] aggregates probability distributions by linear combinations of the local posterior PDF information $F_{\boldsymbol{x}}(\boldsymbol{x} \mid \boldsymbol{Z}_i^r)$ (or appropriate likelihood functions, as per Equation (3.9.11)):

$$F(\boldsymbol{x} \mid \{\boldsymbol{Z}^r\}) = \sum_j w_j F\left(\boldsymbol{x} \mid \boldsymbol{Z}_j^r\right) \tag{3.9.29}$$

where the weights $w_j$ sum to unity and each weight $w_j$ represents a subjective measure of the reliability of the information from sensor $j$. The process can be illustrated as shown in Figure 3.9.6. Bayes' theorem can now be applied to Equation (3.9.29) to obtain a recursive

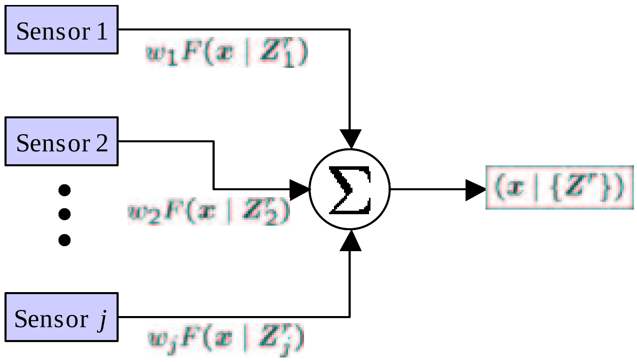


Figure 3.9.6: Multi-Sensor Data Fusion by Linear Opinion Pool

form, which is omitted here for brevity. One of the shortcomings of the linear opinion pool method is its inability to reinforce opinion because the weights are usually unknown except in very specific applications.

The **Independent Opinion Pool** is a product form modification of the linear opinion pool and is defined by the product:

$$F\left(\boldsymbol{x} \mid \{\boldsymbol{Z}^r\}\right) = \alpha \prod_j F\left(\boldsymbol{x} \mid \boldsymbol{Z}_j^r\right) \tag{3.9.30}$$

where $\alpha$ is a normalizing constant. The fusion process in this instance can be illustrated as shown in Figure 3.9.7

This model is widely used since it represents the case when the observations from the individual sensors are essentially independent. However, this is also its weakness, since if the data is correlated at a group of nodes, their opinion is multiplicatively reinforced, which can lead to error propagation in faulty sensor networks. Nevertheless, this technique is appropriate when the prior state space distributions are truly independent and equally likely (as is common in digital communication applications).

To counter the weaknesses of the two common approaches summarized above, a third

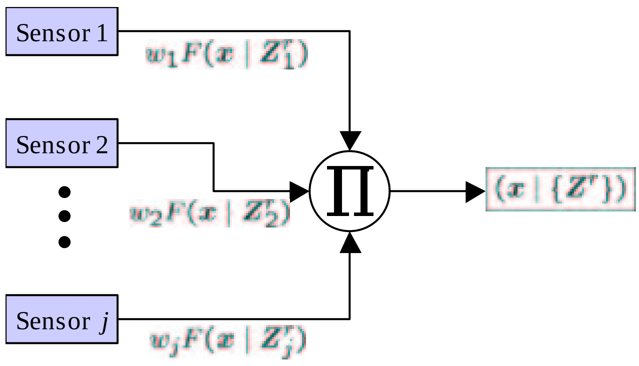


Figure 3.9.7: Multi-Sensor Data Fusion by Independent Opinion Pool

fusion rule is the **Likelihood Opinion Pool**, defined by the following recursive rule.

$$F\left(\boldsymbol{x} \mid \{\boldsymbol{Z}^r\}\right) = \alpha F\left(\boldsymbol{x} \mid \left\{\boldsymbol{Z}^{r-1}\right\}\right) \left[\prod_j \underbrace{F\left(\boldsymbol{z}_j(r) \mid \boldsymbol{x}\right)}_{\text{likelihood}}\right] \tag{3.9.31}$$

The Likelihood Opinion Pool method of data fusion can be illustrated as shown in 3.9.8. The

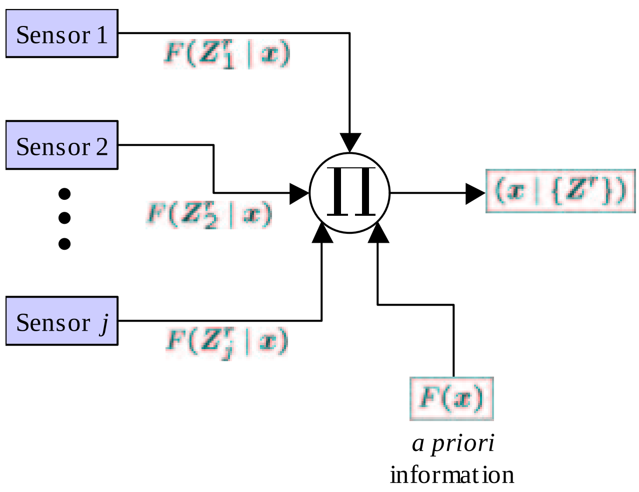


Figure 3.9.8: Multi-Sensor Data Fusion by Likelihood Opinion Pool

likelihood opinion pool technique is essentially a Bayesian update process and is consistent with the recursive process derived in general in Equation (3.9.16). It is interesting to note that a simplified, specific form of this type of information processing occurs in the so called *belief propagation* [28] type of algorithms that are widespread in artificial intelligence and the decoding theory for channel codes. In the exposition above, however, the assumptions and derivations are and explicitly identified and derived, and is thus in a general form that is suitable for application to heterogeneous multi-sensor systems. This provides intuitive insight as to how the probabilistic updates help to reinforce 'opinions' when performing a distributed state space search.

**Information Theoretic Justification of the Bayesian Method**

Probability distributions allow a quantitative description of the observables, the observer, and associated errors. As such, the likelihood functions and distributions contain information about the underlying states that they describe. This approach can be extended further to actually incorporate measures for the information contained in these random variables. In this manner, an information theoretic justification can be obtained for the Likelihood Opinion Pool for multi-sensor data fusion, as discussed in the previous section. Some key concepts from Information Theory [6] are required first.

**Information Measures**

The connections between information theory and distributed detection [44] were briefly surveyed in Section 3.9.2. In this section, some formal information measures are defined to enable an intuitive information theoretic justification of the utility of the Bayesian update method. This approach also provides an insight towards the practical design of algorithms based on the likelihood opinion pool fusion rules that has been discussed earlier.

To build an information theoretic foundation for data fusion, the most useful fundamental metric is the Shannon definition of *Entropy*.

**Definiton 3.9.4. Entropy** is the uncertainty associated with a probability distribution, and is a measure of the descriptive complexity of a PDF [5]. Mathematically:

$$h\{F(\boldsymbol{x})\} \triangleq \mathrm{E}\{-\ln F(\boldsymbol{x})\} \tag{3.9.32}$$

Note that alternative definitions of the concept of information which predate Shannon's formulation, e.g. the *Fisher Information Matrix* [10], are also relevant and useful, but not discussed here further.

Using this definition, an expression for the entropy of the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{Z}^r$ at time $r$ (which is the case of multiple observations from a single sensor) can be expressed as:

$$h(r) \triangleq h\left\{F(\boldsymbol{x} \mid \boldsymbol{Z}^r)\right\} = -\sum F(\boldsymbol{x} \mid \boldsymbol{Z}^r) lnF(\boldsymbol{x} \mid \boldsymbol{Z}^r) \tag{3.9.33}$$

Now, the entropy relationship for Bayes Theorem can be developed as follows:

$$\begin{aligned} \mathrm{E}\left\{-\ln[F(\boldsymbol{x} \mid \boldsymbol{Z}^r)]\right\} &= \mathrm{E}\left\{-\ln[F(\boldsymbol{x} \mid \boldsymbol{Z}^{r-1})]\right\} \\ &- \mathrm{E}\left\{\ln\left[\frac{F(\boldsymbol{z}(r) \mid \boldsymbol{x})}{F(\boldsymbol{z}(r) \mid \boldsymbol{Z}^{r-1})}\right]\right\} \end{aligned} \tag{3.9.34}$$

$$\Longrightarrow h(r) = h(r-1) - \mathrm{E}\left\{\ln\left[\frac{F(\boldsymbol{z}(r) \mid \boldsymbol{x})}{F(\boldsymbol{z}(r) \mid \boldsymbol{Z}^{r-1})}\right]\right\} \tag{3.9.35}$$

This is an alternative form of the result that conditioning with respect to observations reduces entropy (cf. [6]). Using the definition of mutual information, Equation (3.9.34) can be written in an alternative form as shown below.

**Definiton 3.9.5.** For an observation process, *mutual information* at time $r$ is the information about $\boldsymbol{x}$ contained in the observation $\boldsymbol{z}(r)$:

$$\mathrm{I}(\boldsymbol{x}, \boldsymbol{z}(r)) \triangleq \mathrm{E}\left\{\ln\left[\frac{F(\boldsymbol{z}(r) \mid \boldsymbol{x})}{F(\boldsymbol{z}(r))}\right]\right\} \tag{3.9.36}$$

from where

$$h(r) = h(r-1) - \mathrm{I}(r) \tag{3.9.37}$$

which means that the entropy following an observation is reduced by an amount equal to the information inherent in the observation.

The insight to be gained here is that by using the definitions of entropy and mutual information, the recursive Bayes update procedure derived in Equation (3.9.16) can now be seen as an information update procedure:

$$\mathrm{E}\left\{\ln[F(\boldsymbol{x} \mid \boldsymbol{Z}^r)]\right\} = \mathrm{E}\left\{\ln[F(\boldsymbol{x} \mid \boldsymbol{Z}^{r-1})]\right\} + \mathrm{E}\left\{\ln\left[\frac{F(\boldsymbol{z}(r) \mid \boldsymbol{x})}{F(\boldsymbol{z}(r) \mid \boldsymbol{Z}^{r-1})}\right]\right\} \tag{3.9.38}$$

which can be interpreted as [24]:

*posterior information = prior information + observation information.*

The information update equations for the Likelihood Opinion Pool fusion rule thus becomes:

$$\begin{aligned}\mathrm{E}\left\{\ln[F(\boldsymbol{x} \mid \boldsymbol{Z}^r)]\right\} &= \mathrm{E}\left\{\ln[F(\boldsymbol{x} \mid \boldsymbol{Z}^{r-1})]\right\} \\ &+ \sum_j \mathrm{E}\left\{\ln\left[\frac{F(\boldsymbol{z}_j(r) \mid \boldsymbol{x})}{F(\boldsymbol{z}_j(r) \mid \boldsymbol{Z}^{r-1})}\right]\right\}\end{aligned} \tag{3.9.39}$$

The utility of the log-likelihood definition is that the information update steps reduce to simple additions, and are thus amenable to hardware implementation without such problems as overflow and dynamic range scaling.

Thus the Bayesian probabilistic approach is theoretically self-sufficient for providing a unified framework for data fusion in multi-sensor platforms. The information theoretic connection to the Bayesian update makes the approach intuitive, and shows rigorously how the Likelihood Opinion Pool method serves to reduce the ellipsoid uncertainty. This framework answers the question of how to weight or process outputs of diverse sensors, whether they have different sensing modes of signal to noise ratios, without resort to ad hoc criteria. Acoustic, visual, magnetic and other signals can all be combined [17]. Further, since trade-offs in information rate and distortion can be treated using entropies (rate distortion theory [14]) as of course can communication, questions about fundamental limits in sensor networks can now perhaps be systematically explored.

Of course, obvious practical difficulties remain, such as how to determine the uncertainty in measurements, the entropy of sources, and in general how to efficiently convert sensor measurements into entropies.

### 3.9.5 Concluding Remarks

In this section, the approach of using a probabilistic, information processing approach to data fusion in multi-sensor networks was discussed. The Bayesian approach was seen to be the central unifying tool in formulating the key concepts and techniques for decentralized organization of information. Thus, it offers an attractive paradigm for implementation in a wide variety of systems and applications. Further, it allows one to use information theoretic justifications of the fusion algorithms, and also offers preliminary asymptotic analysis of large scale system performance.

The information theoretic formulation makes clear how to combine the outputs of possibly entirely different sensors. Moreover, it allows sensing, signal processing and communication to be viewed in one mathematical framework. This may allow systematic study of many problems involving the cooperative interplay of these elements. This further can lead to the computation of fundamental limits on performance against with practical reduced complexity techniques can be compared.

# Bibliography

[1] J. Balchen *et al.*, *Highly Redundant Sensing in Robotic Systems*, ser. NATO Advanced Science Institutes Series. Springer-Verlag, 1991, vol. 58, ch. Structural Solution of Highly Redundant Sensing in Robotic Systems.

[2] T. Berger *et al.*, "Model distribution in decentralized multi-sensor fusion," in *Proceedings of the American Control Conference (ACC)*, 1991, pp. 2291–2294.

[3] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 405–417, July 1974.

[4] R. S. Blum and S. A. Kassam, "On the asymptotic relative efficiency of distributed detection schemes," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 523–527, March 1995.

[5] D. Catlin, *Estimation, Control and the Discrete Kalman Filter*. Springer-Verlag, 1989.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, New Jersey: Wiley-Interscience, 1991.

[7] I. Csiszar and J. Korner, "Towards a general theory of source networks," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 155–165, 1980.

[8] I. Csiszar and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Sci. Math. Hungar.*, vol. 6, pp. 181–191, 1971.

[9] H. Durrant-Whyte, "Sensor models and multi-sensor integration," *International Journal of Robotics*, vol. 7, no. 6, pp. 97–113, 1988.

[10] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London*, vol. Sec A, no. 222, pp. 309–368, 1922.

[11] A. M. Flynn, "Combining ultra-sonic and infra-red sensors for mobile robot navigation," *International Journal of Robotics Research*, vol. 7, no. 5, pp. 5–14, 1988.

[12] R. Frank, *Understanding Smart Sensors*. Norwood, MA: Artech House, 2000.

[13] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*. Boston, MA: Birkhäuser, 1997, vol. Probability and its Applications.

[14] R. G. Gallager, *Information Theory and Reliable Communications.* New York, NY: John Wiley & Sons, 1968.

[15] E. Gamal and T. M. Cover, "Multiple user information theory," *Proceedings of the IEEE*, vol. 68, pp. 1466–1483, 1980.

[16] T. Garvey *et al.*, "Model distribution in decentralized multi-sensor fusion," *Proceedings of the American Control Conference*, pp. 2291–2294, 1991.

[17] J. W. F. III, M. Wainwright, E. Sudderth, and A. S. Willsky, "Statistical and information-theoretic methods for self-organization and fusion of multimodal, networked sensors," *International Journal of High Performance Computing Applications*, 2001.

[18] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, no. D, pp. 34–35, 1969.

[19] R. Kuc and Siegel, "Physically based simulation model for acoustic sensor robot navigation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 6, pp. 766–778, 1987.

[20] A. D. Kucar, "Mobile radio—an overview," *IEEE Personal Communications Magazine*, pp. 72–85, Nov. 1991.

[21] S. Kullback, *Information Theory and Statistics.* New York: John Wiley & Sons, 1959.

[22] J. J. Leonard, "Directed sonar sensing for mobile robot navigation," Ph.D. dissertation, University of Oxford, 1991.

[23] R. Luo and M. Kay, "Multi-sensor integration and fusion in intelligent systems," *IEEE Transactions on Systems Man and Cybernetics*, vol. 19, no. 5, pp. 901–931, 1989.

[24] J. Manyika and H. Durrant-Whyte, *Data Fusion and Sensor Management*, ser. Ellis Horwood Series in Electrical and Electronic Engineering. West Sussex, England: Ellis Horwood, 1994.

[25] D. Middleton, *Statistical Communication Theory.* McGraw-Hill, 1960.

[26] A. Mitchie and J. K. Aggarwal, "Multiple sensor integration through image processing: A review," *Optical Engineering*, vol. 23, no. 2, 1986.

[27] Y. Nakamura, *Data Fusion, Robotics and Machine Intelligence.* Academic Press, 1992, ch. Geometric Fusion: Minimizing Uncertainty Ellipsoid Volumes.

[28] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1997.

[29] H. V. Poor, *An Introduction to Signal Detection and Estimation.* New York, NY: Springer-Verlag, 1988.

[30] R. Popoli, "The sensor management imperative," in *Multi-Target Multi-Sensor Tracking*, Y. Bar-Shalom, Ed. Artech House, 1992, pp. 325–392.

[31] G. Pottie, "Hierarchical information processing in distributed sensor networks," in *IEEE International Symposium on Information Theory*, Cambridge, MA, August 16-21 1998.

[32] G. Pottie *et al.*, "Wireless sensor networks," in *Information Theory Workshop Proceedings*, Killamey, Ireland, June 22-26 1998.

[33] J. G. Proakis, *Digital Communications.* New York, NY: McGraw-Hill, 2000.

[34] P. Rai-Choudhury, Ed., *MEMS and MOEMS Technology and Applications.* Society of Photo-optical Instrumentation Engineers, 2000.

[35] G. G. Roussas, *A Course in Mathematical Statistics*, 2nd ed. Burlington, MA: Harcourt/Academic Press, 1997.

[36] E. Royer and C.-K. Toh, "A review of current routing protocols for ad hoc wireless networks," *IEEE Personal Communications Magazine*, vol. 6, no. 2, pp. 46–55, April 1999.

[37] J. T. Scheick, *Linear Algebra with Applications.* New York, NY: McGraw-Hill, 1996.

[38] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 279–423, 1948.

[39] S. Singh *et al.*, "Power-aware routing in mobile ad hoc networks," in *Proceedings of the $4^{th}$ Annual IEEE/ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Dallas, TX, 1998, pp. 181–190.

[40] K. Sohrabi and G. Pottie, "Performance of a self-organizing algorithm for wireless ad-hoc sensor networks," *IEEE Vehicular Technology Conference*, Fall 1999.

[41] K. Sohrabi, G. Pottie, *et al.*, "Protocols for self-organization for a wireless sensor network," *IEEE Personal Communications Magazine*, pp. 6–17, Oct. 2000.

[42] M. Stone, "The opinion pool," *The Annals of Statistics*, vol. 32, pp. 1339–1342, 1961.

[43] J. N. Tsitsiklis, "On the complexity of decentralized decision-making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[44] P. K. Varshney, *Distributed Detection and Data Fusion.* New York: Springer-Verlag, 1997.

[45] S. Verdu, *Multiuser Detection.* Cambridge University Press, 1998.

[46] E. Waltz and J. Llinas, *Multi-Sensor Data Fusion.* Artech House, 1991.

[47] K. Yao *et al.*, "Blind beamforming on a randomly distributed sensor array," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, Oct. 1998.