9-1989

Optimal Distributed Decision Fusion

S. C. A. Thomopoulos
Pennsylvania State University - Main Campus

R. Viswanathan
Southern Illinois University Carbondale, viswa@engr.siu.edu

D. K. Bougoulias
Southern Illinois University Carbondale

Correspondence

Optimal Distributed Decision Fusion

The problem of decision fusion in distributed sensor systems is considered. Distributed sensors pass their decisions about the same hypotheses to a fusion center that combines them into a final decision. Assuming that the sensor decisions are independent from each other conditioned on each hypothesis, we provide a general proof that the optimal decision scheme that maximizes the probability of detection at the fusion for fixed false alarm probability consists of a Neyman-Pearson test (or a randomized N-P test) at the fusion and likelihood-ratio tests at the sensors.

I. INTRODUCTION

Systems of distributed sensors monitoring a common volume and passing their decisions into a centralized fusion center which further combines them into a final decision have been receiving a lot of attention in recent years [1]. Such systems are expected to increase the reliability of the detection and be fairly immune to noise interference and to failures. In a number of papers the problem of optimally fusing the decisions from a number of sensors has been considered. Tenney and Sandell [2] have considered the Bayesian detection problem with distributed sensors without considering the design of data fusion algorithms. Sadjadi [3] has considered the problem of hypothesis testing in a distributed environment and has provided a solution in terms of a number of coupled nonlinear equations. The decentralized sequential detection problem has been investigated in [4, 5]. In [6] it was shown that the solution of distributed detection problems is nonpolynomial complete. Chair and Varshney [7] have solved the problem of data fusion when the a-priori probabilities of the tested hypotheses are known and the likelihood-ratio (L-R) test can be implemented at the receiver. Thomopoulos, Viswanathan, and Bougoulias [8, 9] have derived the optimal fusion rule for unknown a-priori probabilities in terms of the Neyman-Pearson (N-P) test.

For the "parallel" sensor topology of Fig. 1, Srinivasan [10] has shown that the globally optimal solution to the fusion problem that maximizes the probability of detection for fixed probability of false alarm when sensors transmit independent, binary decisions to the fusion center, consists of L-R tests

Manuscript received March 31, 1987; revised January 10, 1989.

IEEE Log No. 30105.

This work was supported by the SDIO/IST and managed by the Office of Naval Research under Contract N00014-86-K-0515.

0018-9251/89/0900-0761 $1.00

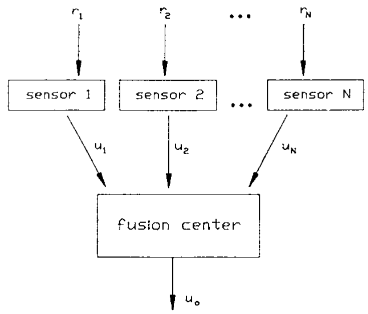

Fig. 1. Distributed sensor fusion. Parallel topology.

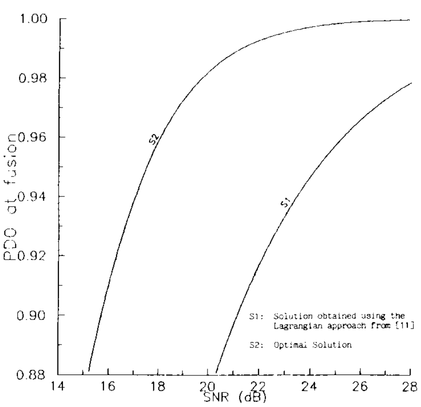

Fig. 2. Example of singularity of Lagrangian approach used in [10] for decision fusion. Three identical sensors in slow-fading Rayleigh channel. Paradigm taken from [11].

at all sensors and a N-P test at the fusion center. This test will be referred to as N-P/L-R hereafter. The proof of the optimality of the N-P/L-R test in [10] is based on the (first-order) Lagrange multipliers methods which does not always yield the optimal solution as it is shown by example in [11]. For the paradigm in [11], the Lagrangian approach fails to yield to optimal solution. Instead, it yields a solution which is by far inferior to the optimal solution, see Fig. 2. A detailed description and analysis of this singular case is given in [11, 12]. A theoretical explanation of the failure of the Lagrange multipliers method can be found in [13, ch. 5, and 14, 15].

In general, if the optimal solution lies on the boundary of the domain of x (as in the decision fusion paradigm in [11]), the Lagrangian formulation fails to guarantee the convexity of the objective function, and thus, the optimality of the solution obtained using the Lagrange multipliers method. In that sense, the proof of optimality of the N-P/L-R test for the parallel sensor topology in [10], which is based on a Lagrangian formulation, is incomplete. We give a complete proof of the optimality of the N-P/L-R test for the distributed decision fusion problem that does not depend on the Lagrangian formulation.

II. OPTIMALITY OF N-P/L-R TEST IN DISTRIBUTED DECISION FUSION

A number of sensors N receive data from a common volume. Sensor k receives data r_k and generates the first stage decision u_k, $k = 1,2,\ldots,N$. The decisions are subsequently transmitted to the fusion center where they are combined into a final decision u_0 about which of the hypotheses is true, Fig. 1. Assuming binary hypothesis testing for simplicity, we use $u_i = 1$ or 0 to designate that sensor i favors hypotheses H_1 or H_0, respectively. In order to derive the globally optimal fusion rule we assume that the received data r_k at the N sensors are statistically independent, conditioned on each hypothesis. This implies that the received decisions at the fusion center are independent conditioned on each hypothesis. Improvement in the performance of conventional diversity schemes is based on the validity of this assumption [16]. Given a desired level of probability of false alarm at the fusion center, $P_{F_0} = \alpha_0$, the test that maximizes the probability of detection P_{D_0} (thus, minimizes the probability of miss $P_{M_0} = 1 - P_{D_0}$) is the N-P test [17, 18]. Because of the comparison to a threshold this test is referred to as a threshold optimal test.

Next, we prove that the optimal solution to the fusion problem involves an N-P test at the fusion center and L-R tests at the sensors.

Let $d(u_1,u_2,\ldots,u_N)$ be the (binary) decision function (rule) at the fusion. Since $d(u_1,u_2,\ldots,u_N)$ is either 0 or 1, and all the possible combinations of decisions $\{u_1,u_2,\ldots,u_N\}$ that the fusion center can receive from the N sensors is 2^N, the set of all possible decision functions contain 2^{2^N} d functions. However, not all these functions d can be threshold optimal as the next Lemma states.

LEMMA 1. *Let the sensors individual decisions u_k be independent from each other conditioned on each hypothesis. Let $P_{F_i} = P(u_i = 1 \mid H_0)$ be the false alarm probability and $P_{D_i} = P(u_i = 1 \mid H_1)$ be the probability of detection at the ith sensors. Assuming, without loss of generality, that for every sensor $P_{D_i} \geq P_{F_i}$, a necessary condition for a fusion function $d(u_1,u_2,\ldots,u_N)$ to be threshold optimal is*

$$d(A_k, U - A_k) = 1 \Rightarrow d(A_n, U - A_n) = 1 \quad \text{if} \quad A_n > A_k \qquad (1)$$

where $U = \{u_1,u_2,\ldots,u_N\}$ denotes the set of the peripheral sensor decisions, A_k is a set of decisions with k sensors favoring hypothesis H_1 *(whereas the complement set of decisions $U - A_k$ favors hypothesis* H_0*), and A_n is any set that contains the decisions from these k sensors. [The symbol ">" is used to indicate "greater than" in the standard multidimensional coordinate-wise sense, i.e., $A_n > A_k$ if and only if $u_{n_i} \geq u_{k_i} \forall i$, $i = 1,2,\ldots,N$, with at least one holding as* *a strict inequality, where $u_{n_i}(u_{k_i})$ indicates the decision of the same ith sensor in the $A_n(A_k)$ decision set.*]

PROOF. Let $P_{F_i} = P(u_i = 1 \mid \mathrm{H}_0)$ be the false alarm probability and $P_{D_i} = P(u_i = 1 \mid \mathrm{H}_1)$ be the probability of detection at the ith sensors. $d(A_k, U - A_k) = 1$ implies that the likelihood ratio

$$\frac{p(A_k, U - A_k \mid \mathrm{H}_1)}{p(A_k, U - A_k \mid \mathrm{H}_0)} = \frac{p(A_k \mid \mathrm{H}_1)p(U - A_k \mid \mathrm{H}_1)}{p(A_k \mid \mathrm{H}_0)p(U - A_k \mid \mathrm{H}_0)} > \lambda_0 \tag{2}$$

which in turn implies that, for $A_n > A_k$,

$$\begin{aligned} &\frac{p(A_n, U - A_n \mid \mathrm{H}_1)}{p(A_n, U - A_n \mid \mathrm{H}_0)} \\ &\quad = \frac{p(A_k \mid \mathrm{H}_1)p(A_n - A_k \mid \mathrm{H}_1)p(U - A_n \mid \mathrm{H}_1)}{p(A_k \mid \mathrm{H}_0)p(A_n - A_k \mid \mathrm{H}_0)p(U - A_n \mid \mathrm{H}_0)} \\ &\quad \geq \frac{p(A_k \mid \mathrm{H}_1)p(U - A_k \mid \mathrm{H}_1)}{p(A_k \mid \mathrm{H}_0)p(U - A_k \mid \mathrm{H}_0)} > \lambda_0 \end{aligned} \tag{3}$$

since, under the assumption that $P_{D_i} \geq P_{F_i}$ for every sensor i,

$$\frac{P(u_i = 1 \mid \mathrm{H}_1)}{P(u_i = 1 \mid \mathrm{H}_0)} = \frac{P_{D_i}}{P_{F_i}} \geq \frac{P(u_i = 0 \mid \mathrm{H}_1)}{P(u_i = 0 \mid \mathrm{H}_0)} = \frac{1 - P_{D_i}}{1 - P_{F_i}}. \tag{4}$$

From (3), it follows that $d(A_n, U - A_n) = 1$.

REMARK 1. Functions that do not satisfy (2) cannot lead to the set of optimal thresholds. A function d that satisfies Lemma 1, is called a monotone increasing function in the context of switching and automata theory, Table I, [19].

REMARK 2. If $P_{D_i} = P_{F_i}$ for all sensors, the L-R at the fusion is degenerated to one, identically for any combination of the peripheral decisions [9]. Hence, **for any likelihood test**, the false alarm probability P_{F_0} and the detection probability P_{D_0} at the fusion are either a) **both one**, if the threshold is less or equal to one, or b) **both zero**, if the threshold is greater than one. In the first case, the fusion rule always favors hypothesis one, independent of the combination of sensor decisions, i.e., $d(U) = 1$ for all Us, which is a monotone increasing function satisfying Lemma 1. In the second case, the fusion rule always favors hypothesis zero, independent of the combination of sensor decisions, i.e., $d(U) = 0$ for all Us, which is a monotone increasing function satisfying Lemma 1.

REMARK 3. If $P_{D_i} \leq P_{F_i}$ for all sensors, the inequality in (3) is reversed, and Lemma 1 still holds with all threshold optimal decisions at the fusion being monotonically increasing functions of the sensor decisions.

REMARK 4. If for some sensors $P_{D_i} \geq P_{F_i}$ while for some others $P_{D_j} \leq P_{F_j}$, Lemma 1 does not hold. However, this is an uninteresting case, for if we wish to maximize the detection probability at the fusion, we would either ignore the sensors for which $P_{D_j} \leq P_{F_j}$, or, randomize their decisions by flipping coins and deciding with probability 1/2 for either one of the two hypotheses.

LEMMA 2. *For any fixed threshold λ_0 and any fixed monotonic function $t(u_1, u_2, \ldots, u_N)$, P_{D_0} is an increasing function of the P_{D_i}s, $i = 1, 2, \ldots, N$.*

PROOF. The decision function that corresponds to the likelihood test at the fusion is contained in the set of monotone functions of N variables. Consider one such monotone increasing decision function $d(u_1, u_2, \ldots, u_N)$. The function d, when expressed in sum of product form in the Boolean sense [19], contains only some of the literals $u_1, \ldots, u_N$ in the uncomplemented form and none of the complemented variables $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_N)$. Since the random variables $u_1, u_2, \ldots, u_N$ are statistically independent, it is possible to compute P_{D_0} knowing the P_{D_i}s [9, eq. (20)–(22)]. Taking partial derivatives of the P_{D_0} w.r.t. P_{D_i}s, one obtains that $(\partial P_{D_0} / \partial P_{D_i}) > 0\ \forall i$, i.e., the desired result. (As an illustration, consider the function $d(u_1, u_2, u_3) = u_1 + u_2 u_3$. For this function $P_{D_0} = P_{D_1} + P_{D_2} P_{D_3} - P_{D_1}(P_{D_2} P_{D_3})$, from which, $(\partial P_{D_0} / \partial P_{D_i}) > 0$, $i = 1, 2, 3$.)

THEOREM 1. *Under the assumption of statistical independence of the sensor decisions conditioned on each hypothesis, the optimal decision fusion rule for the parallel sensor topology consists of an N-P test (or, a randomized N-P test) at the fusion and L-R tests at all sensors.*

PROOF. Given the decisions $u_1, u_2, \ldots, u_N$ at the fusion center, the best fusion rule which achieves maximum P_{D_0} for fixed $P_{F_0} = \alpha_0$ is the N-P test (assuming that the false alarm probability α_0 is realizable by an N-P test at the fusion; the randomized case is treated separately afterwards). Call the best test at the fusion center $t(u_1, \ldots, u_N) \gtrless_{\mathrm{H}_0}^{\mathrm{H}_1} \lambda_0$. From Lemma 1, it follows that the decision function that corresponds to the above test must be one of the monotone increasing functions $d(u_1, u_2, \ldots, u_N)$. Assume that the individual sensors use some test other than the L-R test and are operating with $\{(P_{F_i}, P_{D_i})\ \forall i\}$ such that the condition $P_F = \alpha_0$ is met. From [8, 9] it is seen that P_{F_0} is a function of the P_{F_i}s only, and that P_{D_0} is a function of the P_{D_i}s only. Furthermore, from Lemma 2, P_{D_0} is a monotonic increasing function of the P_{D_i}s. Therefore, the L-R tests at the sensors which operate with $(P_{F_i}^* = P_{F_i}, P_{D_i}^*)$ lead to the best performance at the fusion, since in this case, the achieved $P_{D_0}^*$ is greater than or equal to P_{D_0} that can be achieved with any other test at the sensors.

If the false alarm probability α_0 is not achievable by an N-P test, a randomized N-P maximizes the

TABLE I
Number Of Monotone Increasing Functions And Percentage of Reduction

Number of Sensors N	Number of Monotone Functions	Number of all Possible 2^{2^N} Functions	Percentage Reduction
1	3	4	25
2	6	16	62.5
3	20	256	92.19
4	168	65,536	99.74
5	7,581	4.2949673×10^9	99.99982
6	7,828,354	1.8446744×10^{19}	100

TABLE II
Total Number Of Functions Searched For The Set Of Optimal Thresholds

Number of Sensors N	L_N (is number of Monotone Functions -2)	Total Number of Functions R_N	Percentage Reduction
1	1	1	0.00
2	4	2	50.00
3	18	9	50.00
4	166	114	31.13
5	7,579	6,894	9.03
6	7,828,352	7,786,338	0.54

probability of detection at the fusion for the given false alarm probability. Let the best randomized N-P test at the fusion center be $\mathbf{t}(u_1,\ldots,u_N) \gtrless_{H_0}^{H_1} \lambda_0$ w.p. p, resulting in false alarm probability P_{F_0}, and $\bar{t}(u_1,\ldots,u_N) \gtrless_{H_0}^{H_1} \bar{\lambda}_0$ w.p. $1-p$, resulting in false alarm probability $\bar{P}_{F_0}$. The thresholds λ_0 and $\bar{\lambda}_0$ are chosen so that the total false alarm at the fusion

$$P_{F_0} = pP_{F_0} + (1-p)\bar{P}_{F_0} = \alpha_0. \qquad (5)$$

Thus, the corresponding detection probability at the fusion

$$P_{D_0} = pP_{D_0} + (1-p)\bar{P}_{D_0}. \qquad (6)$$

Since the probability p is fixed from the constraint (5), the detection probability in (6) is maximized if each one of the P_{D_0} and $\bar{P}_{D_0}$ is maximized. But, according to the part of the proof in the nonrandomized N-P test above, each one of these two detection probabilities is maximized if an L-R test is used at the sensors. Hence, the Theorem is also proven for the randomized N-P/L-R test.

A precise characterization of the set of fusion functions that satisfy Theorem 1, indicated as R_N in Table II, can be found in [12].

III. CONCLUSIONS

A general proof that the optimal fusion rule for the distributed detection problem of Fig. 1 involves an N-P test (or a randomized N-P test) at the fusion and L-R tests at all sensors has been provided. The proof does not suffer from the weaknesses of the Lagrange-multipliers-based proof in [10].

S. C. A. THOMOPOULOS*
R. VISWANATHAN
D. K. BOUGOULIAS
Dep't. of Electrical Engineering
Southern Illinois University
Carbondale, IL 62901

***Currently with:**
Dep't. of Electrical Engr.
The Pennsylvania State Univ.
University Park, PA 16802

REFERENCES

[1] Conte, E., D'Addio, E., Farina, A., and Longo, M. (1983)
Multistatic radar detection: synthesis and comparison of optimum and suboptimum receivers.
IEEE Proc. F, Commun., Radar & Signal Process, 1983.

[2] Tenney, R. R., and Sandell, N. R., Jr. (1981)
Detection with distributed sensors.
IEEE Transactions on Aerospace and Electronic Systems, **AES-17** (July 1981), 501–510.

[3] Sadjadi, F. A. (1986)
Hypothesis testing in a distributed environment.
IEEE Transactions on Aerospace and Electronic Systems, **AES-22** (Mar. 1986), 134–137.

[4] Teneketzis, D., and Varaiya, P. (1984)
The decentralized quickest detection problem.
IEEE Transactions on Automatic Control, **AC-29**, 7 (July 1984), 641–644.

[5] Teneketzis, D. (1982)
The decentralized Wald problem.
In *Proceedings of the IEEE 1982 International Large-Scale Systems Symposium*, Virginia Beach, VA, Oct. 1982, 423–430.

[6] Tsitsiklis, J., and Athans, M. (1985)
On the complexity of distributed decision problems.
IEEE Transactions on Automatic Control, **AC-30**, 5 (May 1985), 440–446.

[7] Chair, Z., and Varshney, P. K. (1986)
Optimal data fusion in multiple sensor detection systems.
IEEE Transactions on Aerospace and Electronic Systems, **AES-22**, 1 (Jan. 1986), 98–101.

[8] Thomopoulos, S. C. A., Viswanathan, R., and Bougoulias, D. K. (1986)
Optimal decision fusion in multiple sensor systems.
In *Proceedings of the 24th Allerton Conference*, Monticello, IL, Oct. 1–3, 1986, 984–993.

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

[10] Srinivasan, R. (1986)
Distributed radar detection theory.
IEE Proceedings, **133**, Pt. F, 1 (Feb. 1986), 55–60.

[11] Viswanathan, R., and Thomopoulos, S. C. A. (1987)
Distributed data fusion.
Technical report, TR-SIU-DEE-87-4, Department of Electrical Engineering, Southern Illinois University, Carbondale, Apr. 1987.

[12] Thomopoulos, S. C. A., Viswanathan, R., and Bougoulias, D. K. (1988)
Optimal and suboptimal distributed decision fusion. 22nd Annual Conference on Information Sciences and Systems, Princeton University, NJ, Mar. 16–18, 1988, 886–890.

[13] Hestenes, M. R. (1975)
Optimization Theory: The Finite Dimensional Case. New York: Wiley, 1975.

[14] Thomopoulos, S. C. A., and Viswanathan, R. (1988)
Optimal and suboptimal distributed decision fusion. Technical Report, TR-SIU-DEE-87-5, Department of Electrical Engineering, Southern Illinois University, Carbondale, Jan. 1988.

[15] Thomopoulos, S. C. A., Bougoulias, D. K., and Zhang, L. (1988)
Optimal and suboptimal distributed decision fusion. Presented at SPIE 1988 Technical Symposium on Optics, Electro-Optics, and Sensors, Apr. 4–8, 1988, Orlando, FL.

[16] Viswanathan, R., Thomopoulos, S. C. A., and Tumuluri, R. (1988)
Optimal serial distributed decision fusion. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-24**, 4 (July 1988), 366–376.

[17] Van Trees, H. L. (1968)
Detection Estimation and Modulation Theory, Vol I. New York: Wiley, 1968.

[18] DiFranco, J. V., and Rubin, W. L. (1980)
Radar Detection. Dedham, MA: Artech House, 1980.

[19] Harrison, M. A. (1965)
Introduction to Switching and Automata Theory. New York: McGraw-Hill, 1965.

Decoding Techniques in State Estimation for Dynamic Systems With Past Histories

States of discrete dynamic systems with past histories are first quantized and then estimated by using both the Viterbi decoding algorithm and a stack sequential decoding algorithm. State estimation with a stack sequential decoding algorithm is faster and more practical than the state estimation with the Viterbi decoding algorithm, even though the estimates obtained by the Viterbi decoding algorithm are superior to the estimates by a stack sequential decoding algorithm.

I. INTRODUCTION

Researchers have been dealing with recursive state estimation of dynamic systems with a first-order memory since Kalman's original work [5]. As a result, many estimation schemes have been proposed [5–11], and these schemes have been also applied for practical systems [12]. These estimation schemes are referred to as the classical estimation schemes. Dynamic models of the classical estimation schemes, which are said to be the classical dynamic models, must be linear functions of a white disturbance noise and (additive) observation noise, and they must also have a first-order memory. Well-known optimum state estimates have been presented for linear dynamic models with white Gaussian noise. However, optimum state estimates cannot, in general, be given for nonlinear dynamic models except for some special cases. An example of these cases is the classical nonlinear discrete dynamic models with discrete state values and white Gaussian noise. The states of these models can be optimally estimated (in the mean-square sense) by recursively computing the conditional density of a state given the observations, and then finding the conditional mean of this state [9]. States of nonlinear dynamic models are, in general, estimated by linearizing nonlinear models by a Taylor series expansion [6, 9]. Hence, nonlinear functions of nonlinear models must be smooth enough for a Taylor series expansion. Linearization errors may sometimes cause state estimates to diverge from the actual state values [13].

Recently, Demirbaş [1], and Demirbaş and Leondes [2, 3] have considered state estimation of dynamic models with a first-order memory, which are more general than the models of the classical estimation schemes. These dynamic models can be nonlinear functions of the states, disturbance noise, and observation noise. The resulting estimation schemes are based upon the decoding techniques of information theory. These schemes have been also applied for practical systems [4]. These schemes do not require any model linearizations. Therefore, the state estimate divergence caused by model linearization errors are prevented with these schemes. Thus, these schemes are superior to the classical estimation schemes, such as the extended Kalman filter, for highly nonlinear dynamic systems [4].

States of dynamic models with a higher order memory (i.e., with a memory of order which is greater than one) could be estimated by first representing these dynamic models by higher dimensional dynamic models with a first-order memory, and then using an estimation scheme cited above. But this increases the implementation complexity of state estimation.

Here, states of dynamic models with a higher order memory are estimated by using both a stack sequential decoding algorithm and the Viterbi decoding algorithm (VDA), without higher dimensional dynamic system representation. This results in memory reduction for state estimate implementation.

Manuscript received July 12, 1988; revised January 30, 1989.

IEEE Log No. 30104.

0018-9251/89/0900-0765 $1.00 © 1989 IEEE

II. PROBLEM STATEMENT

We treat the state estimation of dynamic systems with past histories (i.e., an Mth-order memory),