Distributed Multitarget Classification in Wireless Sensor Networks

Jayesh H. Kotecha, *Member, IEEE*, Vinod Ramachandran, and Akbar M. Sayeed, *Senior Member, IEEE*

***Abstract*—We study distributed strategies for classification of multiple targets in a wireless sensor network. The maximum number of targets is known *a priori* but the actual number of distinct targets present in any given event is assumed unknown. The target signals are modeled as zero-mean Gaussian processes with distinct temporal power spectral densities, and it is assumed that the noise-corrupted node measurements are spatially independent. The proposed classifiers have a simple distributed architecture: local hard decisions from each node are communicated over noisy links to a manager node which optimally fuses them to make the final decision. A natural strategy for local hard decisions is to use the optimal local classifier. A key problem with the optimal local classifier is that the number of hypotheses increases exponentially with the maximum number of targets. We propose two suboptimal (mixture density and Gaussian) local classifiers that are based on a natural but coarser repartitioning of the hypothesis space, resulting in linear complexity with the number of targets. We show that exponentially decreasing probability of error with the number of nodes can be guaranteed with an arbitrarily small but nonvanishing communication power per node. Numerical results based on real data demonstrate the remarkable practical advantage of decision fusion: an acceptably small probability of error can be attained by fusing a moderate number of unreliable local decisions. Furthermore, the performance of the suboptimal mixture density classifier is comparable to that of the optimal local classifier, making it an attractive choice in practice.**

***Index Terms*—Decision fusion, error exponents, hypothesis testing.**

I. Introduction

WIRELESS sensor networks promise an unprecedented opportunity to monitor the physical world via cheap wireless nodes that can sense the environment in multiple modalities, including acoustic, seismic, and infrared [1], [2]. Detection and classification of objects in the sensor field is an important application of sensor networks. Due to a variety of factors, such as measurement noise and statistical variability in target signals, collaborative processing of multiple node measurements is necessary for reliable decision making.

In practical implementation, collaborative processing of sensor measurements is typically coordinated by a manager node (see, e.g., [3]). Given the limited communication ability of sensor nodes (due to limited transmit power), a key goal in developing collaborative signal processing (CSP) algorithms is to transmit the least amount of data from the sensing nodes to the manager node. With this goal in mind, in this paper we extend our earlier work on CSP algorithms for single target classification [4], [5] to the more challenging multiple target setting.

The proposed distributed classifiers are based on modeling each target as a point source whose temporal signal characteristics can be modeled as a zero-mean stationary Gaussian process [4]–[6]. Each target generates a Gaussian space–time signal field whose statistical characteristics have a profound impact on classifier performance. In particular, the region of interest can be divided into spatial coherence regions (SCRs) over which the spatial signal field remains strongly correlated. The size of the SCRs is inversely proportional to the target signal bandwidth: high-bandwidth targets induce smaller SCRs, whereas low-bandwidth targets induce larger SCRs [5], [6]. An important property of the SCRs is that the spatial signal in distinct SCRs is approximately uncorrelated (independent in the Gaussian case). Thus, the number of SCRs in the query region determines the number of *independent* spatial measurements that can be collected at any given time. The sensor nodes sample the spatio-temporal signal field generated by targets.

From a sensing perspective, there are two main sources of error in distributed decision making: 1) sensor measurement noise and 2) the inherent statistical variability in the target signal. The above modeling of the space–time target signal field in terms of SCRs suggests a natural structure on CSP algorithms to mitigate the two sources of error [4]–[6]. First, since all nodes within each SCR sense a highly correlated target signal, the node measurements in each SCR can be aggregated to improve the effective measurement signal-to-noise ratio (SNR). Second, the independent node measurements from distinct SCRs can be combined to reduce the impact of target signal variability. Furthermore, since the node measurements in distinct SCRs are approximately independent, local decisions can be first formed in each SCR, and then the lower dimensional decisions can be communicated to the manager node.

Consider a network query regarding the classification of multiple distinct targets in a region of interest. We assume that all targets are distinct and the maximum number (M) of targets is known *a priori*. However, the actual number of targets present in a given event is unknown. Thus, the multitarget classification problem corresponds to a N-ary hypothesis testing problem with $N = 2^M$ hypotheses corresponding to all possibilities for the presence or absence of each target. Thus, the complexity of optimal classifiers (centralized or decentralized) increases exponentially with the number of targets M. To circumvent this

Manuscript received December 1, 2003; revised July 30, 2004. This work was supported by the Defense Advanced Research Projects Agency (DARPA) SensIT Program under Grant F30602-00-2-0555 and in part by the National Science Foundation (NSF) under ITR Grant CCR-0113385.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA (e-mail: jkotecha@ece.wisc.edu; vinod@cae.wisc.edu; akbar@engr.wisc.edu).

Digital Object Identifier 10.1109/JSAC.2005.843539

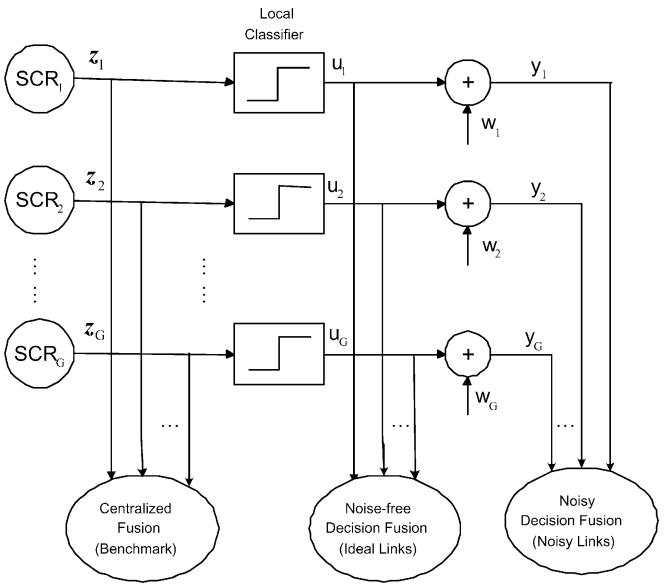


Fig. 1. Generic architecture for multitarget classification. All classifiers operate on G i.i.d. feature vectors $\{z_1, \ldots, z_G\}$ collected in distinct SCRs. Local hard decisions from the SCRs are communicated via ideal or noisy links to a manager node that makes the final decision.

exponential complexity, we propose two suboptimal [mixture Gaussian (MG) and single Gaussian (SG)] distributed classifiers based on a natural repartitioning of the hypothesis space: the suboptimal classifiers perform M (rather than $N = 2^M$) tests to determine the presence or absence of targets.

The basic architecture of the proposed distributed classifiers is illustrated in Fig. 1. We assume that there are G SCRs in the query region. In a given snapshot, an N_o-dimensional feature vector $\boldsymbol{z}_k$ is collected in the kth SCR. The optimal centralized classifier that directly operates on $\boldsymbol{z}_k$ provides a upper bound on the performance of any classifier (centralized or decentralized). A local hard decision u_k about the N hypotheses is made in each SCR based on the local feature vector $\boldsymbol{z}_k$. We consider three local classifiers for generating $\{u_k\}$: the optimal local classifier, and two suboptimal (MG and SG) classifiers based on the repartitioned hypothesis space. The local decisions are then communicated over ideal (noise-free) or noisy links to the manager node that optimally fuses them to make the final decision. Thus, we consider three distributed classification schemes based on the three local classifiers (optimal, MG, SG). Performance analysis of the classifiers shows that the probability of error decreases exponentially with the number of independent node measurements, as long as each node communicates with a nonvanishing power. Numerical results based on real data demonstrate the remarkable advantage of decision fusion in practice: an acceptably small probability of error can be attained by fusing a moderate number of fairly unreliable local decisions. Furthermore, the performance of the suboptimal mixture density classifier is comparable to that of the optimal local classifier, making it an attractive choice in practice.

In Section II, we formulate the multitarget classification problem. In Section III, we analyze the performance of the optimal centralized classifier that serves as a benchmark and also facilitates analysis of the distributed classifiers. Section IV discusses the three local classifiers for generating the decisions in each SCR. In Section V, we present the structure of the final distributed classifier at the manager node that optimally fuses the decisions from the SCRs. Section VI presents numerical results and Section VII provides concluding remarks.

TABLE I
HYPOTHESIS SPACE FOR $M = 2$ TARGETS

H_j	b_2	b_1	π_j
H_0	0	0	$q_2 q_1$
H_1	0	1	$q_2(1-q_1)$
H_2	1	0	$(1-q_2)q_1$
H_3	1	1	$(1-q_2)(1-q_1)$

II. PROBLEM FORMULATION

The problem of interest is the classification of multiple target/objects in a query region of the sensor field. In a practical scenario, a query for target classification will usually be preceded by a query for target detection. Target detection can be accomplished reliably with distributed energy detectors (see, e.g., [3]). However, in this paper, we include the null hypothesis (no targets present) in the formulation. We assume that *at most* M distinct targets are present in the query region during the duration of interest. However, the actual number of targets present is unknown. Consequently, there are $N = 2^M$ possible hypotheses, denoted by H_j, $j = 0, \ldots, N-1$, corresponding to all possibilities for the presence or absence of each target. The probability of the mth target being present is denoted by $1 - q_m$ and is independent of other targets. Let $b_m(j)$ denote the presence $(b_m(j) = 1)$ or absence $(b_m(j) = 0)$ of the mth target under H_j. The prior probability of H_j is given by

$$\pi_j = \prod_{m=1}^{M} [b_m(j)(1-q_m) + (1-b_m(j))q_m] \tag{1}$$

where $j = 0, \ldots, N-1$, and $b_M(j)b_{M-1}(j)\cdots b_1(j)$ is the binary representation of j, H_0 corresponds to no target present, whereas H_{N-1} corresponds to all targets present. See Table I for an example with $M = 2$.

The final decision about the hypotheses is made at the manager node based on G independent identically distributed (i.i.d.) effective feature vectors $\{\boldsymbol{z}_k\}$ collected in G distinct SCRs. Each $\boldsymbol{z}_k$ is of dimension N_o and its signal component corresponding to mth target is modeled as a zero-mean complex Gaussian vector with covariance matrix $\boldsymbol{\Sigma}_m$. The energy in each target is assumed to be the same; that is $\mathrm{tr}(\boldsymbol{\Sigma}_m) = \sigma_s^2$ for all m. It follows that the signal corresponding to each H_j is also Gaussian. The multitarget classification problem can be stated as as N-ary hypothesis testing problem $(j = 0, \ldots, N-1)$

$$\begin{aligned} &H_j : \boldsymbol{z}_k = \boldsymbol{s}_k + \boldsymbol{n}_k, \quad k = 1, \ldots, G \\ &\boldsymbol{s}_k \sim \mathcal{CN}(\boldsymbol{0}, \overline{\boldsymbol{\Sigma}}_j), \quad \overline{\boldsymbol{\Sigma}}_j = \sum_{m=1}^{M} b_m(j)\boldsymbol{\Sigma}_m \\ &\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}). \end{aligned} \tag{2}$$

Under H_j, the probability density function (pdf) of the feature vector in the kth SCR is

$$p_j(\boldsymbol{z}_k) = \frac{1}{\pi^{N_o}|\widetilde{\boldsymbol{\Sigma}}_j|} e^{-\boldsymbol{z}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k}, \qquad \widetilde{\boldsymbol{\Sigma}}_j = \overline{\boldsymbol{\Sigma}}_j + \sigma^2 \boldsymbol{I}.$$

III. Optimal Centralized Classifier

In this section, we review the optimal *centralized* classifier for the hypothesis testing problem (2) whose probability of error (P_e) serves as an ultimate lower bound for the P_e of any distributed classifier. The optimal classifier is given by

$$C(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) = \arg \max_{j=0,\ldots,N-1} p_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G)\pi_j$$
$$p_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) = p(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G | H_j) = \prod_{k=1}^{G} p_j(\boldsymbol{z}_k)$$

due to the conditional independence of measurements $\boldsymbol{z}_k$. In terms of log-likelihoods

$$C(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) = \arg \min_{j=0,\ldots,N-1} l_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) \tag{3}$$
$$l_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) = -\frac{1}{G} \log p_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G)\pi_j$$
$$= -\frac{1}{G} \sum_{k=1}^{G} \log p_j(\boldsymbol{z}_k) - \frac{1}{G} \log \pi_j. \tag{4}$$

Ignoring constants that do not depend on the class, l_j takes the form

$$l_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) = \log|\widetilde{\boldsymbol{\Sigma}}_j| + \frac{1}{G} \sum_{k=1}^{G} \boldsymbol{z}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k - \frac{1}{G} \log \pi_j.$$

Note that implementation of the optimal centralized classifier requires that the kth SCR communicates the local log-likelihood functions for the N hypotheses, $\left\{\boldsymbol{z}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k,\ j = 0, \ldots, N-1\right\}$, to the manager node. The classifier at the manager node then makes the decision according to (3).

A. Performance of the Optimal Centralized Classifier

We quantify classifier performance in terms of the average probability of error

$$P_e(G) = \sum_{m=0}^{N-1} P_{e,m}(G)\pi_m \tag{5}$$
$$P_{e,m}(G) = P(l_j < l_m \text{ for some } j \neq m \,|H_m) \tag{6}$$

where $P_{e,m}(G)$ is the conditional error probability under H_m. Computing $P_{e,m}$ is complicated in general but we can bound it using the union bound [7]

$$P_{e,m}(G) \leq \sum_{j=0, j\neq m}^{N-1} P(l_j < l_m | H_m) \tag{7}$$
$$P_e(G) \leq \sum_{m=0}^{N-1} \sum_{j=0, j\neq m}^{N-1} P(l_j < l_m | H_m)\pi_m. \tag{8}$$

For each pair (j, m), the pairwise error probability (PEP), $\mathrm{PEP}_{jm}(G) = P(l_j < l_m | H_m)$ depends on a decision statistic that is a weighted sum of NG χ_2^2 random variables [4]. The pdf and distribution function of the statistic can be computed in closed-form but take on tedious expressions [7]. Chernoff bounding techniques can be used to obtain tight bounds for the PEPs. We state some well-known results (see, e.g., [7] and [8]) in the context of our set up. Let $E_m[\cdot]$ denote the expectation under H_m.

Proposition 1 (Chernoff Bounds): For any $0 \leq \theta \leq 1$ and for all $G \geq 1$

$$\mathrm{PEP}_{jm}(G) \leq \left(\frac{\pi_j}{\pi_m}\right)^{\theta} e^{\mu_{jm}(\theta)G},$$
$$\mu_{jm}(\theta) = \log E_m \left[\left(\frac{p_j(\boldsymbol{Z})}{p_m(\boldsymbol{Z})}\right)^{\theta}\right] \leq 0. \tag{9}$$

For any G, the tightest Chernoff bound for $\mathrm{PEP}_{jm}(G)$ can be obtained by minimizing (9) over θ

$$\mathrm{PEP}_{jm}(G) \leq \left(\frac{\pi_j}{\pi_m}\right)^{\theta^*(G)} e^{-D^*_{jm}(G)G}$$
$$D^*_{jm}(G) = -\mu_{jm}(\theta^*(G))$$
$$\theta^*(G) = \arg \min_{0\leq\theta\leq 1} \left[\frac{\theta}{G} \log\left(\frac{\pi_j}{\pi_m}\right) + \mu_{jm}(\theta)\right]. \tag{10}$$

The tightest error exponent (Chernoff information) in the limit of large G is given by

$$D^*_{jm} = -\lim_{G\to\infty} \mu_{jm}(\theta^*(G)) = -\min_{0\leq\theta\leq 1} \mu_{jm}(\theta). \tag{11}$$

A simple nontrivial exponent for all G is the Bhattacharyya bound which corresponds to $\theta = 1/2$.

For the optimal classifier, $\mu_{jm}(\theta)$ is given by [5]

$$\mu_{jm}(\theta) = \theta \log\left|\widetilde{\boldsymbol{\Sigma}}_m \widetilde{\boldsymbol{\Sigma}}_j^{-1}\right| - \log\left|(1-\theta)\boldsymbol{I} + \theta\widetilde{\boldsymbol{\Sigma}}_m \widetilde{\boldsymbol{\Sigma}}_j^{-1}\right|$$

where the above expression holds for all $\theta \geq 0$ for which $(1-\theta)\boldsymbol{I} + \theta\widetilde{\boldsymbol{\Sigma}}_m \widetilde{\boldsymbol{\Sigma}}_j^{-1}$ is positive definite. The minimization in (11) can be performed numerically.

Using (8) and (10), we get the following bound:

$$P_e(G) \leq \sum_{m=0}^{N-1} \sum_{j=0, j\neq m}^{N-1} A_{jm}(G) e^{-D^*_{jm}(G)G} \tag{12}$$

where $A_{jm}(G) = \pi_j^{\theta^*(G)} \pi_m^{1-\theta^*(G)}$. Since all PEP_{jm}'s decay exponentially to zero with G, so does $P_e(G)$. In particular, for a given SNR, the decay of $P_e(G)$ will be dominated by the smallest (worst) error exponent.

Asymptotic performance as $G \to \infty$. Note from (4) that by the law of large numbers, under H_m, we have

$$\lim_{G\to\infty} l_j(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G) = -E_m[\log p_j(\boldsymbol{Z})]$$
$$= -\lim_{G\to\infty} \frac{1}{G} \sum_{k=1}^{G} \log p_j(\boldsymbol{z}_k)$$
$$- \lim_{G\to\infty} \frac{1}{G} \log \pi_j$$

where we note that the second term becomes zero, implying that the effect of the prior probabilities on the decision vanishes in the limit. Furthermore

$$-E_m[\log p_j(\boldsymbol{Z})] = D(p_m\|p_j) + h_m(\boldsymbol{Z}) \tag{13}$$

where $D(p_m\|p_j)$ is the Kullback–Leibler (K-L) distance between the pdfs p_j and p_m and $h_m(\boldsymbol{Z})$ is the differential entropy of $\boldsymbol{Z}$ under H_m [9]

$$\begin{aligned} D(p_m\|p_j) &= E_m\left[\log\left(\frac{p_m(\boldsymbol{Z})}{p_j(\boldsymbol{Z})}\right)\right] \\ &= \log\left(\frac{|\widetilde{\boldsymbol{\Sigma}}_j|}{|\widetilde{\boldsymbol{\Sigma}}_m|}\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{\Sigma}}_m - \boldsymbol{I}\right) \\ h_m(\boldsymbol{Z}) &= -E_m[\log p_m(\boldsymbol{Z})] = \log\left((\pi e)^{N_o}|\widetilde{\boldsymbol{\Sigma}}_m|\right). \end{aligned}$$

Thus, under H_m, we get

$$\lim_{G\to\infty} l_j(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_G) - l_m(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_G) = D(p_m\|p_j)$$

where $D(p_m\|p_j) \geq 0$ with equality if and only if $p_m = p_j$. This implies that

$$\lim_{G\to\infty} P(l_j < l_m|H_m) = 0$$

if and only if $D(p_m\|p_j) > 0$ for all $j \neq m$. It follows from (5) and (6) that $P_e(G)$ goes to zero as $G \to \infty$ if and only if all the pairwise K-L distances are positive. From Proposition 1, we know that each PEP_{jm} goes to zero exponentially with G if the corresponding exponent is positive. It can be shown that the Chernoff exponent for a particular PEP_{jm} is strictly positive if and only if the corresponding K-L distances are strictly positive [5]. The behavior of P_e of the optimal centralized classifier is summarized in the following result.

Proposition 2: $P_e(G)$ decays exponentially to zero with G

$$\lim_{G\to\infty} \frac{1}{G}\log P_e(G) \leq -D^*_{\min} \tag{14}$$

if and only if all pairwise K-L distances are strictly positive

$$D(p_m\|p_j) > 0 \qquad \forall\, j,m,\ j \neq m \tag{15}$$

where $D^*_{\min} = \min\{D^*_{jm}\}$ is the smallest pairwise Chernoff information defined in (11).

The condition (15) in the above result would be satisfied, in principle, even at arbitrarily low measurement SNR, provided the targets have distinct statistics. However, the values of the K-L distances would get smaller for lower measurement SNRs.

IV. Local Classifiers

In the optimal centralized classifier, the kth SCR sends N local log-likelihood values, $\left\{\boldsymbol{z}_k^H\widetilde{\boldsymbol{\Sigma}}_j^{-1}\boldsymbol{z}_k : j = 0,\ldots,N-1\right\}$, to the manager node (see Fig. 1). While communication of real-valued likelihoods puts much less communication burden on the network as compared with *data fusion* in which the feature vectors $\{\boldsymbol{z}_k\}$ are communicated, it is attractive to reduce the burden even further. A natural quantization strategy is to compute local *hard decisions* in each SCR. The local decisions, denoted as $\{u_k\}$, from all SCRs are communicated to the manager node which makes the final optimal[1] decision. The local classifier in the kth SCR makes a decision using only its local measurement $\boldsymbol{z}_k$ according to a given decision rule $f : \mathcal{C}^{N_o} \to \{0,1,\ldots,N-1\}$

$$u_k = f(\boldsymbol{z}_k), \qquad k = 1,\ldots,G. \tag{16}$$

Since all $\{\boldsymbol{z}_k\}$ are i.i.d., so are $\{u_k\}$. The decision statistics are characterized by the probability mass functions (pmfs), $\{p_m[j]\}$, of the decision U under different hypotheses

$$p_m[j] = P(U = j|H_m) \qquad j,m = 0,\ldots,N-1. \tag{17}$$

Since $\{u_k\}$ are i.i.d., the pmfs $\{p_m[j]\}$ are identical for all SCRs.

In this section, we discuss three local classifiers (and corresponding $\{p_m[j]\}$). This first one is the optimal classifier which has exponential complexity in the number of targets M. The other two are suboptimal classifiers, based on a repartitioning of the hypothesis space, which have linear complexity in M.

A. Optimal Local Classifier

The optimal Bayesian local classifier in the kth SCR is given by

$$u_k = \arg\max_{j=0,\ldots,N-1} p_j(\boldsymbol{z}_k)\pi_j, \qquad k = 1,\ldots,G \tag{18}$$

and is illustrated in Fig. 2(a). The pmfs of U under different hypotheses are characterized by the following probabilities:

$$p_m[j] = P\left(p_j(\boldsymbol{z}_k)\pi_j \geq p_l(\boldsymbol{z}_k)\pi_l\ \forall\, l \neq j\ |H_m\right) \tag{19}$$

where $j,m = 0,\ldots,N-1$. The complexity of the optimal classifier increases exponentially with M since $N = 2^M$.

B. Suboptimal Local Classifiers

In the suboptimal classifiers, M tests are conducted, one for each target, to determine the presence or absence of the target.

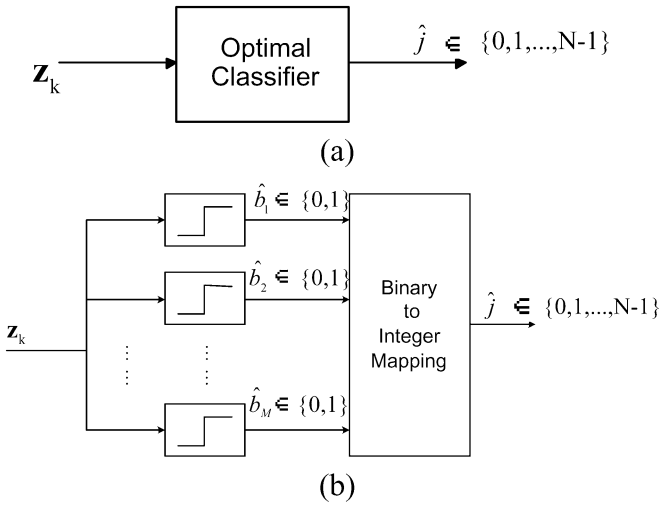


Fig. 2. (a) Optimal local classifier. (b) Suboptimal local classifiers.

[1]Optimal, given the statistics of decisions $\{u_k\}$ and the nature of the communication link.

The test for mth target partitions the hypothesis space into two sets, $\widetilde{H}_m$ (target present) and $\widetilde{H}_m^c$ (target absent). Let

$$S_m = \{j \in \{0,\dots,N-1\} : b_m(j) = 1\}$$
$$S_m^c = \{j \in \{0,\dots,N-1\} : b_m(j) = 0\}$$

where recall that $b_M(j)b_{M-1}(j)\dots b_1(j)$ is the binary representation of j. Then, the two hypotheses for the mth test are

$$\widetilde{H}_m = \bigcup_{j\in S_m} H_j$$
$$\widetilde{H}_m^c = \bigcup_{j\in S_m^c} H_j = \left\{\bigcup_{j=0}^{N-1} H_j\right\} - \widetilde{H}_m.$$

In the example in Table I, $\widetilde{H}_1 = H_1 \cup H_3$, $\widetilde{H}_1^c = H_0 \cup H_2$, $\widetilde{H}_2 = H_2 \cup H_3$, and $\widetilde{H}_2^c = H_0 \cup H_1$.

Under $\widetilde{H}_m$ and $\widetilde{H}_m^c$, $\boldsymbol{z}_k$ is distributed as a weighted sum of Gaussians [mixture Gaussian (MG)]

$$p(\boldsymbol{z}_k|\widetilde{H}_m) = \frac{1}{1-q_m}\sum_{i\in S_m}\pi_i p_i(\boldsymbol{z}_k) = \frac{1}{1-q_m}\sum_{i\in S_m}\pi_i \frac{1}{\pi^{N_o}|\widetilde{\boldsymbol{\Sigma}}_i|} e^{-\boldsymbol{z}_k^H\widetilde{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{z}_k} \quad (20)$$

$$p(\boldsymbol{z}_k|\widetilde{H}_m^c) = \frac{1}{q_m}\sum_{i\in S_m^c}\pi_i p_i(\boldsymbol{z}_k) = \frac{1}{q_m}\sum_{i\in S_m^c}\pi_i \frac{1}{\pi^{N_o}|\widetilde{\boldsymbol{\Sigma}}_i|} e^{-\boldsymbol{z}_k^H\widetilde{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{z}_k}. \quad (21)$$

We now present two local classifiers based on this repartitioned hypothesis space: the MG classifier which is the optimal classifier for the repartitioned space, and the SG classifier that approximates the mixture densities with a Gaussian density. Essentially, under H_j, the mth test estimates the value of $b_m(j)$. Let $\widehat{b}_m \in \{0,1\}$ denote the value of the mth test. Let $f_{\text{bin}-\text{int}}$ denote the binary-to-integer mapping: $f_{\text{bin}-\text{int}}(b_M, b_{M-1},\dots,b_1) = \sum_{m=1}^{M} b_m 2^{m-1}$. Then, $u_k(\boldsymbol{z}_k) = f_{\text{bin}-\text{int}}(\widehat{b}_M(\boldsymbol{z}_k), \widehat{b}(\boldsymbol{z}_k),\dots,\widehat{b}_1(\boldsymbol{z}_k))$ is the integer representation of the binary decisions $\widehat{b}_M\ \widehat{b}_{M-1}\cdots\widehat{b}_1$ [see Fig. 2(b)]. In both the MG and the SG classifiers, the pmfs of the i.i.d. local decisions in each SCR are characterized by the following probabilities:

$$p_m[j] = P(\hat{b}_M = b_M(j),\dots,\hat{b}_1 = b_1(j)|H_m)$$

where $j, m = 0,\dots,N-1$.

1) Mixture Gaussian (MG) Classifier: This is the optimal classifier for the repartitioned hypothesis space. For $m = 1,\dots,M$, let $\widehat{b}_m(\boldsymbol{z}_k)$ denote the value of the mth binary test between $\widetilde{H}_m$ and $\widetilde{H}_m^c$ in the kth SCR

$$(1-q_m)p(\boldsymbol{z}_k|\widetilde{H}_m) \gtrless_{\widehat{b}_m=0}^{\widehat{b}_m=1} q_m p\left(\boldsymbol{z}_k|\widetilde{H}_m^c\right)$$
$$\Updownarrow$$
$$\sum_{i\in S_m}\pi_i p(\boldsymbol{z}_k|H_i) \gtrless_{\widehat{b}_m=0}^{\widehat{b}_m=1} \sum_{i\in S_m^c}\pi_i p(\boldsymbol{z}_k|H_i). \quad (22)$$

Consider the example with $M = 2$. It is easy to show that for $m = 1$, the test in (22) can be written as be

$$q_2\left[(1-q_1)\mathcal{CN}(\boldsymbol{0},\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_1) - q_1\mathcal{CN}(\boldsymbol{0},\sigma^2\boldsymbol{I})\right] + (1-q_2)\left[(1-q_1)\mathcal{CN}(\boldsymbol{0},\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2) - q_1\mathcal{CN}(\boldsymbol{0},\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_2)\right] \gtrless_{\widehat{b}_1=0}^{\widehat{b}_1=1} 0. \quad (23)$$

A similar expression can be written for $m = 2$. Observe that the above test is a weighted sum of two tests. The first expression in square brackets is the Bayes test for detecting target 1 given that target 2 is absent, which is weighted by the probability of target 2 being absent. Similarly, the second expression in square brackets is the Bayes test for detecting target 1 given that target two is present, which is weighted by $(1-q_2)$.

2) Single Gaussian (SG) Classifier: The SG classifier is obtained by approximating the distributions in (20) and (21) by SGs

$$\hat{p}(\boldsymbol{z}_k|\widetilde{H}_m) = \frac{1}{\pi^{N_o}|\widehat{\boldsymbol{\Sigma}}_m|} e^{-\boldsymbol{z}_k^H\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{z}_k}$$
$$\hat{p}\left(\boldsymbol{z}_k|\widetilde{H}_m^c\right) = \frac{1}{\pi^{N_o}|\widehat{\boldsymbol{\Sigma}}_{-m}|} e^{-\boldsymbol{z}_k^H\widehat{\boldsymbol{\Sigma}}_{-m}^{-1}\boldsymbol{z}_k}$$

where $\widehat{\boldsymbol{\Sigma}}_m = \sum_{i\in S_m}\pi_i\widetilde{\boldsymbol{\Sigma}}_m$ and $\widehat{\boldsymbol{\Sigma}}_{-m} = \sum_{i\in S_m^c}\pi_i\widetilde{\boldsymbol{\Sigma}}_m$. For $m = 1,\dots,M$, the mth test in the kth SCR is given by

$$(1-q_m)\hat{p}(\boldsymbol{z}_k|\widetilde{H}_m) \gtrless_{\widehat{b}_m=0}^{\widehat{b}_m=1} q_m\hat{p}\left(\boldsymbol{z}_k|\widetilde{H}_m^c\right).$$

V. Distributed Classifiers: Fusion of Local Decisions

In the previous section, we discussed three local classifiers for generating the local hard decisions $\{u_k\}$ in the different SCRs: the optimal local classifier, the suboptimal MG classifier and the suboptimal SG classifier. In this section, we discuss the structure of the corresponding final distributed classifier at the manager node that optimally fuses the local decisions. As illustrated in Fig. 1, we consider ideal (noise-free) communication links as well as noisy links corrupted by additive white Gaussian noise (AWGN). The performance of the final classifier at the manager node can be characterized in a unified fashion for all three local classifiers. For ideal communication links, the performance of the final classifier is determined by the pmfs of the local decisions, $\{p_j[m] : j, m = 0,\dots,N-1\}$. For noisy communication links, the performance is governed by noisy pdfs under different hypotheses induced by the pmfs of local decisions and the AWGN channel.

Let $P_{e,\text{opt,ideal}}$, $P_{e,\text{mg,ideal}}$ and $P_{e,\text{sg,ideal}}$ denote the P_e of the final classifier under noise-free communication corresponding to the optimal local classifier, suboptimal MG classifier, and the suboptimal SG classifier, respectively. Similarly, let $P_{e,\text{opt,noisy}}$, $P_{e,\text{mg,noisy}}$ and $P_{e,\text{sg,noisy}}$ denote the P_e under noisy communication links. As a general trend, we expect $P_{e,\text{opt}} \le P_{e,\text{mg}} \le P_{e,\text{sg}}$ and $P_{e,\text{ideal}} \le P_{e,\text{noisy}}$.

A. Decision Fusion With Ideal Communication Links

With ideal communication links, the final classifier at the manager node is given by

$$C_{\text{ideal}}(u_1,\ldots,u_G) = \arg \min_{j=0,\ldots,N-1} l_{j,\text{ideal}}[u_1,\ldots,u_G]$$

$$l_{j,\text{ideal}}[u_1,\ldots,u_G] = -\frac{1}{G}\sum_{k=1}^{G} \log p_j[u_k] - \frac{1}{G}\log \pi_j. \quad (24)$$

The above expressions apply to all three local classifiers; the only difference is that they induce different decision pmf's $\{p_j[m] = P(U = m|H_j)\}$.

1) Performance of Ideal Decision Fusion: While the exact calculation of $P_{e,\text{ideal}}$ is difficult, it can be bounded and analyzed asymptotically analogous to the centralized classifier.

Bound on $P_{e,\text{ideal}}$. As in the centralized classifier, $P_{e,\text{ideal}}(G)$ decays exponentially with G and is bounded analogous to (9) with $\mu_{jm,\text{ideal}}(\theta)$ given by

$$\mu_{jm,\text{ideal}}(\theta) = \log E_m\left[\frac{p_j^{\theta}[U]}{p_m^{\theta}[U]}\right] = \log \sum_{i=0}^{N-1} p_j^{\theta}[i] p_m^{1-\theta}[i].$$

Then, the PEPs and $P_{e,\text{ideal}}$ can be bounded analogous to (10) and (12).

Asymptotic performance as $G \to \infty$. Despite the fact that the local hard decisions can be quite unreliable, the final hard decision fusion classifier can still attain perfect performance as $G \to \infty$. From (24) and the law of large numbers, under H_m, we have

$$\lim_{G\to\infty} l_{j,\text{ideal}}[u_1,\ldots,u_G] = -E_m[\log p_j[U]] = D(p_m\|p_j) + H_m(U)$$

where $D(p_m\|p_j)$ is the K-L distance between the pmf's p_m and p_j and $H_m(U)$ is the entropy of the hard decision under H_m [9]

$$D(p_m\|p_j) = \sum_{i=0}^{N-1} p_m[i]\log\left(\frac{p_m[i]}{p_j[i]}\right)$$

$$H_m(U) = -\sum_{i=0}^{N-1} p_m[i]\log p_m[i].$$

It follows that $\lim_{G\to\infty} P_{e,\text{ideal}} = 0$ if and only if $D(p_m\|p_j) > 0, \forall\, m \neq j$. The following result summarizes the behavior of $P_{e,\text{ideal}}$ for all three local classifiers.

Proposition 3: $P_{e,\text{ideal}}(G)$ decays exponentially to zero with G

$$\lim_{G\to\infty} \frac{1}{G}\log P_{e,\text{ideal}}(G) \le -D^*_{\min,\text{ideal}}$$

if and only if all the pairwise K-L distances between the decision pmfs are strictly positive

$$D(p_j\|p_m) > 0, \qquad \forall\, j \neq m$$

where $D^*_{\min,\text{ideal}} = \min\{D^*_{jm,\text{ideal}}\}$ is the smallest pairwise Chernoff information corresponding to the underlying local classifier (optimal, MG, or SG).

Note that for a given measurement SNR, the error exponent for optimal ideal decision fusion will be smaller compared with the centralized classifier due to local *hard* decisions: the pairwise K-L distances between the pmfs of the optimal local classifier will be smaller than those between the pdfs in the centralized classifier. Amongst the three local classifiers, the K-L distances would generally decrease from optimal to suboptimal MG to suboptimal SG classifiers.

B. Decision Fusion With Noisy Communication Links

In this section, we discuss fusion of local SCR decisions via noisy communication links, as illustrated in Fig. 1. We assume that each SCR has a dedicated communication link to the manager node.[2] Each SCR sends a scaled version of its local decision u_k in (16) over an AWGN link

$$y_k = \alpha u_k + w_k, \qquad k = 1,\ldots,G \quad (25)$$

where y_k denotes the received signal at the manager node from the kth SCR and $\{w_k\}$ are i.i.d. $\mathcal{N}(0,\sigma_w^2)$ (real Gaussian noise). Note that since $\{u_k\}$ are i.i.d., so are $\{y_k\}$. Without loss of generality, assume that N is odd and define $\widetilde{N} = (N-1)/2$. We assume that each SCR sends a symmetrized version of its decision to use minimum power: $u_k \in \{-\widetilde{N},\ldots,\widetilde{N}\}$. Given this simple communication scheme, the optimal decentralized classifier at the manager node is given by

$$C_{\text{noisy}}(\boldsymbol{y}) = \arg\min_j l_{j,\text{noisy}}(\boldsymbol{y})$$

$$l_{j,\text{noisy}}(\boldsymbol{y}) = -\frac{1}{G}\sum_{k=1}^{G}\log p_{j,\text{noisy}}(y_k)$$

$$p_{j,\text{noisy}}(y) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\sum_{i=-\widetilde{N}}^{\widetilde{N}} e^{-(y-\alpha i)^2/2\sigma_w^2} p_j[i]. \quad (26)$$

1) Performance of Noisy Decision Fusion: The exact calculation of $P_{e,\text{noisy}}$ is most complicated in this case; however, it can be bounded and analyzed asymptotically as in the centralized classifier and ideal fusion.

Bound on $P_{e,\text{noisy}}$. As in the centralized classifier, Chernoff and Bhattacharya bounds for the PEPs can be obtained. In particular, in this case, $\mu_{jm,\text{noisy}}(\theta)$ is given by

$$\mu_{jm,\text{noisy}}(\theta) = \log E_m\left[\frac{p_{j,\text{noisy}}^{\theta}(Y)}{p_{m,\text{noisy}}^{\theta}(Y)}\right]$$

which can be computed numerically.

Asymptotic performance as $G \to \infty$. For sufficiently large measurement and communication SNRs, we again expect $P_{e,\text{noisy}} \to 0$ as $G \to \infty$. From (26), we have under H_m

$$\lim_{G\to\infty} l_{j,\text{noisy}}(\boldsymbol{y}) = -E_m[\log p_{j,\text{noisy}}(Y)] = D(p_{m,\text{noisy}}\|p_{j,\text{noisy}}) + h_m(Y).$$

Hence, $\lim_{G\to\infty} P_{e,\text{noisy}} = 0$ if and only if $D(p_{m,\text{noisy}}\|p_{j,\text{noisy}}) > 0, \forall\, m \neq j$. The following summarizes the performance of noisy decision fusion.

[2]Note that this requires large bandwidth or large latency at the manager node in the limit of large G.

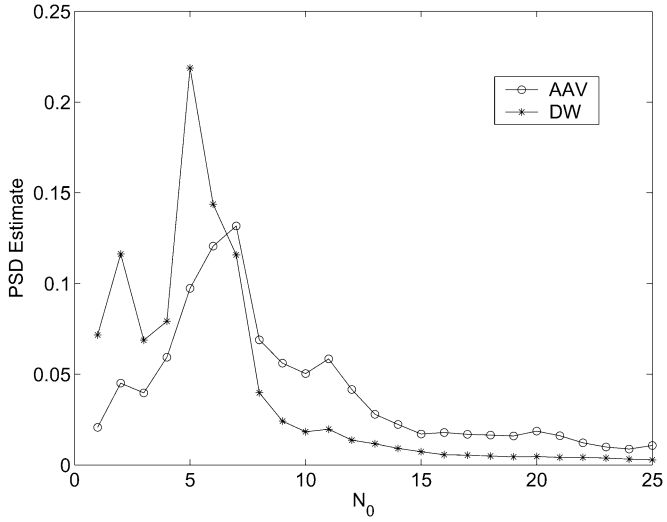


Fig. 3. Covariance matrix eigenvalues (PSD estimates) of acoustic measurements for the two target vehicles.

Proposition 4: $P_{e,\text{noisy}}$ decays exponentially to zero as $G \to \infty$

$$\lim_{G\to\infty} \frac{1}{G} P_{e,\text{noisy}} \leq -D^*_{\text{min,noisy}}$$

if and only if all pairwise K-L distances between noisy pdfs in (26) are strictly positive

$$D(p_{j,\text{noisy}} \| p_{m,\text{noisy}}) > 0, \quad \forall\, m \neq j$$

where $D^*_{\text{min,noisy}} = \min\{D^*_{jm,\text{noisy}}\}$ is the smallest pairwise Chernoff information for noisy decision fusion corresponding to the underlying local classifier (optimal, MG, or SG).

Note that $D^*_{\text{min,noisy}} > 0$ will be guaranteed, in principle, even at arbitrarily low communication SNRs, as long as $D^*_{\text{min,ideal}} > 0$. However, for any local classifier, we expect the K-L distances in this case to be smaller than those for ideal decision fusion due to communication noise.

VI. Numerical Results

Numerical results presented in this section are based on real acoustic data collected during the DARPA SensIT program for two ($M = 2$) targets: amphibious assault vehicle (AAV) and dragon wagon (DW). PSD values at $N_o = 25$ frequencies (within a 2 kHz bandwidth) were estimated using data collected at multiple nodes. The PSD estimates are plotted in Fig. 3. The PSD values define the diagonal covariance matrices (in the Fourier domain) for the two targets: $\mathbf{\Sigma}_m = \mathbf{\Lambda}_m, m = 1{,}2$. Thus, the simulated multitarget classification problem corresponds to N=4 hypotheses. Under H_j, $j = 0{,}1{,}2{,}3$, the $N_o = 25$-dimensional feature vector $\boldsymbol{z}_k$ in the kth SCR was simulated as

$$\boldsymbol{z}_k = \overline{\mathbf{\Lambda}}_j^{1/2} \boldsymbol{v}_k + \boldsymbol{n}_k, \qquad k = 1, \ldots, G$$

where $\{\boldsymbol{v}_k\}$ are i.i.d. $\mathcal{CN}(\mathbf{0}, \boldsymbol{I})$, $\overline{\mathbf{\Lambda}}_j = b_1(j)\mathbf{\Lambda}_1 + b_2(j)\mathbf{\Lambda}_2$, and $\{\boldsymbol{n}_k\}$ are i.i.d. $\mathcal{CN}\left(\mathbf{0}, \sigma_n^2 \boldsymbol{I}\right)$.

The three proposed distributed classifiers, corresponding to the three local classifiers (optimal, MG, and SG), are compared with the optimal centralized classifier under ideal (noise-free) and noisy communication links. The optimal centralized classifier serves as a benchmark for the proposed classifiers. The P_e's for the different classifiers are plotted as a function of G for various measurement and communication SNRs which are defined as

$$\text{SNR}_{\text{meas}} = 10 \log_{10} \left(\frac{\sigma_s^2}{N_o \sigma_n^2} \right)$$

$$\text{SNR}_{\text{comm}} = 10 \log_{10} \left(\sum_i \frac{(\alpha i)^2 p[i]}{\sigma_w^2} \right)$$

where $p[i] = P(U = i) = \sum_j p_j[i]\pi_j$. We assume equal prior probabilities ($q_1 = q_2 = 1/2$) and equal signal energies $\left(\sigma_s^2 = \text{tr}(\mathbf{\Lambda}_1) = \text{tr}(\mathbf{\Lambda}_2)\right)$ for the two targets. The $P_{e,\text{ideal}}$'s for noise-free links were estimated via Monte Carlo simulation using 10 000 independent sets of G measurements assuming equal prior probabilities, i.e., $q_1 = 1/2$ and $q_2 = 1/2$. The pmf's $\{p_m[i]\}$ of decisions for the three local classifiers were also estimated via this Monte Carlo simulation. The measurements for noisy decision fusion were simulated using (25) and the pmfs for hard decisions, and the $P_{e,\text{noisy}}$'s were estimated using 50 000 independent sets of realizations.

Figs. 4–6 plot $P_e(G)$ for $\text{SNR}_{\text{meas}} = 0$, 4, and 10 dB, respectively. For each SNR_{meas}, noisy links with $\text{SNR}_{\text{comm}} = 0$ and 10 dB are considered. In all figures, $P_e(G)$ of the centralized classifier serves as a lower bound. Note that even with ideal communication links, the distributed classifiers incur a loss in error exponent (slope of P_e) compared with the centralized classifier due to local *hard* decisions. As expected, the P_e of all classifiers improves with SNR_{meas} and $P_{e,\text{noisy}}$ is higher than $P_{e,\text{ideal}}$ but approaches it as SNR_{comm} increases. The performance of the SG classifier is worse than both the MG and optimal classifiers. Most importantly, in general, the suboptimal MG classifier performs nearly as well as the optimal classifier for all considered values of SNR_{meas} and SNR_{comm}.

As the figures show, P_e decays exponentially with G for all classifiers, albeit with different rates (exponents). This demonstrates an important practical advantage of multiple independent measurements in sensor networks: we can attain reliable classification performance by fusing a relatively moderate number of much less reliable independent local decisions. The rates for the optimal and the MG classifier are greater than those for the SG classifier. These differences in performance can be attributed to the differences in the pair-wise K-L distances shown in Table II for the different classifiers. We can infer the following general trends from the values of K-L distances.

- The K-L distances increase with SNR_{meas} for all classifiers.
- For a given SNR_{meas}, the K-L distances decrease from the centralized classifier to the distributed optimal to MG to SG classifiers.
- For any distributed classifier, the K-L distances are lower for noisy links than those for ideal links, and they increase with SNR_{comm}.

However, these general observations are violated in some instances. This may be partly attributed to numerical inaccuracies in Monte Carlo estimates.

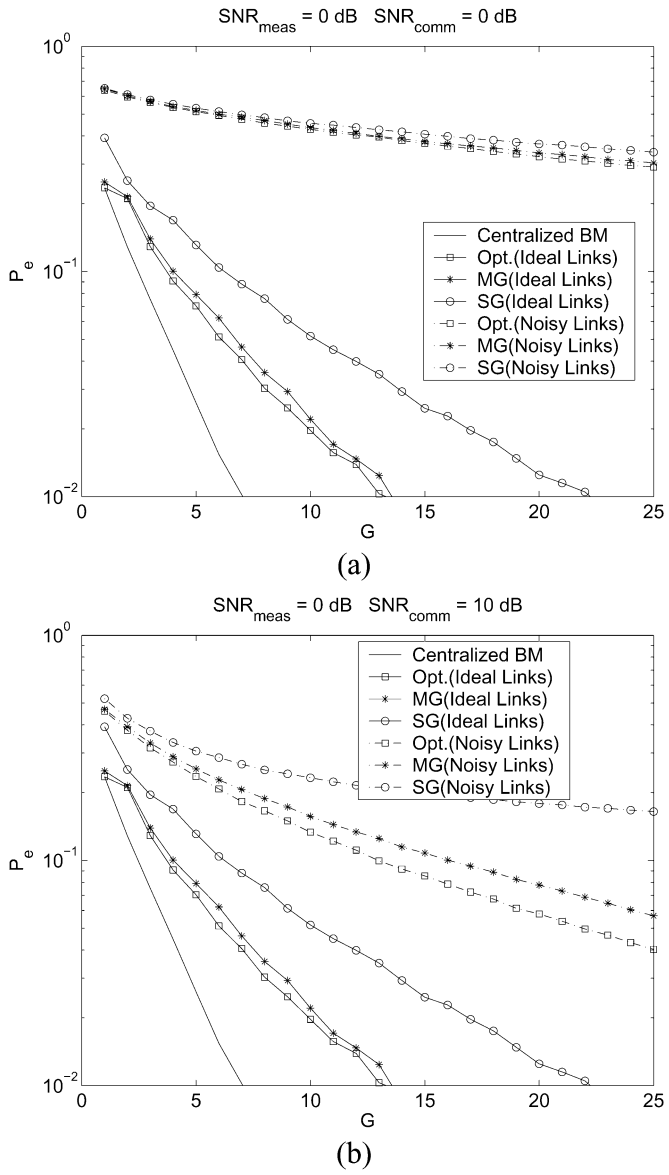


Fig. 4. $P_e(G)$ at $\mathrm{SNR}_{\mathrm{meas}} = 0$ dB for the distributed classifiers and the centralized benchmark (BM) under ideal and noisy communication links. (a) $\mathrm{SNR}_{\mathrm{comm}} = 0$ dB. (b) $\mathrm{SNR}_{\mathrm{comm}} = 10$ dB.

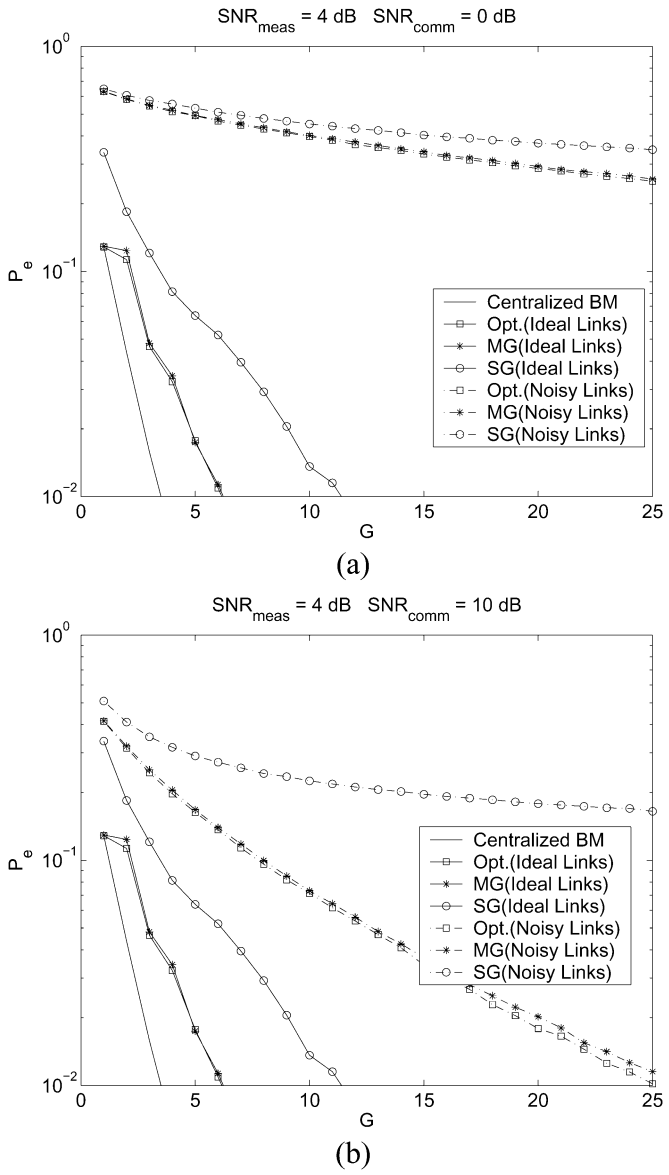


Fig. 5. $P_e(G)$ at $\mathrm{SNR}_{\mathrm{meas}} = 4$ dB for the distributed classifiers and the centralized benchmark (BM) under ideal and noisy communication links. (a) $\mathrm{SNR}_{\mathrm{comm}} = 0$ dB. (b) $\mathrm{SNR}_{\mathrm{comm}} = 10$ dB.

Fig. 7 compares the union bounds for $P_{e,\mathrm{ideal}}$ with the actual simulated $P_{e,\mathrm{ideal}}$ for $\mathrm{SNR}_{\mathrm{meas}}$ of -4 and 10 dB. Observe that the bounds match the error exponent fairly well but exhibit an offset. Furthermore, the bounds get tighter at higher $\mathrm{SNR}_{\mathrm{meas}}$. We note that tighter approximations for $P_e(G)$ with negligible offsets can be obtained via accurate estimates of the constants multiplying the exponentials [5].[3]

Finally, we also performed simulations for $M = 3$ targets (the third vehicle was a Humvee). Plots of P_e are shown in Fig. 8 for $\mathrm{SNR}_{\mathrm{meas}}$ of 10 dB and $\mathrm{SNR}_{\mathrm{comm}}$ of 10 dB. As evident, there is some loss in performance of the MG classifier compared with the optimal classifier, relative to $M = 2$ targets. This is natural since the signal space is getting more crowded due to the larger number of hypotheses, which reduces the pairwise K-L distances. In general, the feature vector dimension N_o would have to be increased appropriately as M increases.

[3]The constants decay as $\mathcal{O}(1/\sqrt{G})$.

VII. Discussion and Conclusion

We have taken a first step in attacking the challenging problem of distributed multitarget classification in wireless sensor networks. A key problem with multitarget classification is that the number of hypotheses increases exponentially with the number of targets. To circumvent this exponential complexity, we proposed two suboptimal distributed classifiers based on a natural repartitioning of the hypothesis space that results in linear complexity. Our results show that the suboptimal MG classifier, that is optimal for the repartitioned space, delivers performance comparable to the optimal distributed classifier.

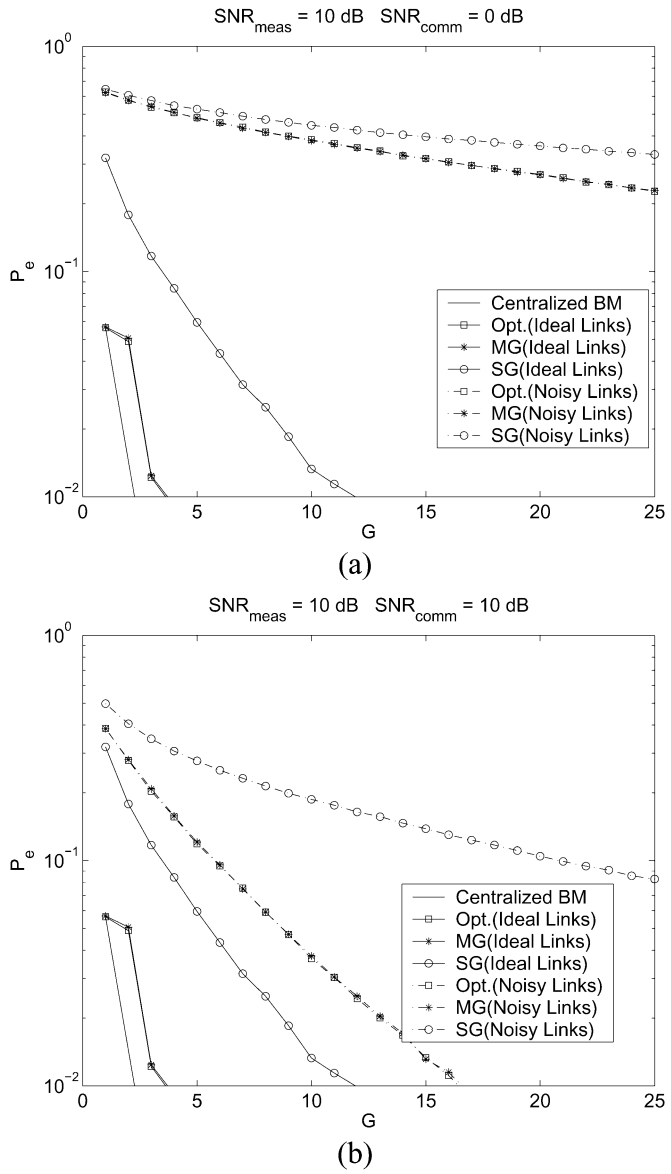


Fig. 6. $P_e(G)$ at $\mathrm{SNR}_{\mathrm{meas}} = 10$ dB for the distributed classifiers and the centralized benchmark (BM) under ideal and noisy communication links. (a) $\mathrm{SNR}_{\mathrm{comm}} = 0$ dB. (b) $\mathrm{SNR}_{\mathrm{comm}} = 10$ dB.

TABLE II
PAIRWISE K-L DISTANCES FOR $\mathrm{SNR}_{\mathrm{meas}} = -4$ dB (10 dB). THE FIRST FOUR ENTRIES IN EACH CELL CORRESPOND TO THE OPTIMAL CENTRALIZED, THE DISTRIBUTED OPTIMAL, THE MG, AND THE SG CLASSIFIERS WITH IDEAL LINKS. THE NEXT THREE ENTRIES CORRESPOND TO THE DISTRIBUTED OPTIMAL, THE MG, AND THE SG CLASSIFIERS WITH NOISY LINKS AT $\mathrm{SNR}_{\mathrm{comm}} = 10$ dB. THE FINAL THREE VALUES CORRESPOND TO THE DISTRIBUTED OPTIMAL, THE MG, AND THE SG CLASSIFIERS WITH NOISY LINKS AT $\mathrm{SNR}_{\mathrm{comm}} = 0$ dB

Measurement SNR = -4 dB (10 dB)				
m/j	No Vehicle	AAV	DW	AAV and DW
No Vehicle	0 (0)	1.48 (32.30)	1.74 (25.30)	3.71 (42.8)
	0 (0)	0.99 (13.81)	1.32 (13.81)	2.68 (13.81)
	0 (0)	0.93 (13.81)	1.19 (13.81)	2.63 (13.81)
	0 (0)	0.70 (4.39)	0.89 (1.60)	2.26 (13.81)
	0 (0)	0.46 (1.10)	0.84 (3.39)	1.62 (3.37)
	0 (0)	0.49 (1.10)	0.83 (3.40)	1.77 (3.38)
	0 (0)	0.43 (1.83)	0.67 (1.07)	1.69 (5.15)
	0 (0)	0.09 (0.13)	0.18 (0.39)	0.38 (0.74)
	0 (0)	0.11 (0.13)	0.19 (0.39)	0.39 (0.74)
	0 (0)	0.09 (0.37)	0.15 (0.21)	0.37 (0.60)
AAV	2.19 (196.28)	0 (0)	0.47 (8.62)	0.84 (3.14)
	1.42 (13.45)	0 (0)	0.25 (3.18)	0.62 (1.98)
	1.33 (13.44)	0 (0)	0.27 (3.21)	0.59 (1.98)
	1.36 (13.02)	0 (0)	0.19 (1.29)	0.60 (1.02)
	0.60 (1.45)	0 (0)	0.08 (0.69)	0.42 (1.43)
	0.74 (1.46)	0 (0)	0.06 (0.69)	0.41 (1.43)
	0.82 (4.29)	0 (0)	0.03 (0.11)	0.44 (0.33)
	0.10 (0.14)	0 (0)	0.01 (0.07)	0.10 (0.26)
	0.13 (0.14)	0 (0)	0.01 (0.07)	0.08 (0.26)
	0.11 (0.41)	0 (0)	0.01 (0.02)	0.09 (0.02)
DW	3.03 (206.58)	0.52 (6.44)	0 (0)	0.71 (7.71)
	1.91 (13.64)	0.25 (3.32)	0 (0)	0.44 (4.60)
	1.77 (13.63)	0.27 (3.32)	0 (0)	0.40 (4.68)
	1.84 (9.92)	0.16 (3.13)	0 (0)	0.41 (8.42)
	1.09 (3.95)	0.08 (0.57)	0 (0)	0.17 (0.60)
	1.22 (3.96)	0.06 (0.57)	0 (0)	0.16 (0.60)
	1.25 (2.65)	0.03 (0.12)	0 (0)	0.23 (0.96)
	0.20 (0.39)	0.01 (0.06)	0 (0)	0.04 (0.07)
	0.21 (0.40)	0.01 (0.07)	0 (0)	0.03 (0.07)
	0.18 (0.24)	0.01 (0.02)	0 (0)	0.05 (0.09)
AAV and DW	7.38 (434.81)	1.19 (6.08)	0.89 (15.26)	0 (0)
	4.23 (13.47)	0.62 (2.14)	0.45 (4.39)	0 (0)
	3.09 (13.48)	0.47 (2.13)	0.33 (4.05)	0 (0)
	3.51 (13.80)	0.48 (0.28)	0.32 (1.19)	0 (0)
	2.09 (8.12)	0.41 (1.49)	0.17 (0.76)	0 (0)
	2.18 (8.09)	0.34 (1.48)	0.14 (0.74)	0 (0)
	2.51 (5.98)	0.36 (0.15)	0.18 (0.43)	0 (0)
	0.41 (0.80)	0.09 (0.26)	0.04 (0.07)	0 (0)
	0.41 (0.80)	0.08 (0.26)	0.03 (0.07)	0 (0)
	0.40 (0.60)	0.08 (0.02)	0.04 (0.08)	0 (0)

The proposed classifiers have a simple distributed architecture that naturally limits the communication burden: the network is partitioned into SCRs and local hard decisions from each SCR are communicated to the fusion center that makes the final decision. For given target covariance matrices, the P_e of the distributed classifiers is controlled by: 1) the $\mathrm{SNR}_{\mathrm{meas}}$ at the sensors; 2) the $\mathrm{SNR}_{\mathrm{comm}}$ of the AWGN links; and 3) the number (G) of independent node decisions from distinct SCRs. The effective $\mathrm{SNR}_{\mathrm{meas}}$ can be improved by densely sampling the nodes within each SCR and averaging their measurements, which requires local node cooperation within each SCR [5]. The effective $\mathrm{SNR}_{\mathrm{comm}}$ can be improved by increasing the transmit power or using coded communication between the sensor nodes and the manager node (our simulation were based on uncoded communication). For given $\mathrm{SNR}_{\mathrm{meas}}$ and $\mathrm{SNR}_{\mathrm{comm}}$, the final P_e at the manager node can be reduced exponentially by fusing multiple (G) independent node measurements from distinct SCRs. Thus, for a target P_e, the results of this paper can be used to determine the desirable combination of sampling density, transmit power and channel coding, and the size of the region over which sensor measurements are taken (G).

Without latency constraints, $\mathrm{SNR}_{\mathrm{meas}}$ is the limiting factor on performance: $P_{e,\mathrm{noisy}}$ can be made arbitrarily close to $P_{e,\mathrm{ideal}}$ by using progressively powerful channel coding. Furthermore, this can be achieved, in principle, with an arbitrarily small but nonvanishing power per node since each node has to communicate information at a vanishing rate. Our numerical results show that even at $\mathrm{SNR}_{\mathrm{meas}} = 0$ dB, relatively moderate values of G (10–20) are sufficient to attain $P_{e,\mathrm{ideal}} \approx 0.01$ with the suboptimal MG classifier.

An important related issue is the impact of signal path loss in sensing. We assumed that all node measurements from different SCRs were i.i.d. In practice they will be independent but will not be identically distributed: nodes farther form the target will exhibit poorer $\mathrm{SNR}_{\mathrm{meas}}$. Thus, we expect an optimal radius around each target within which nodes measurements should be

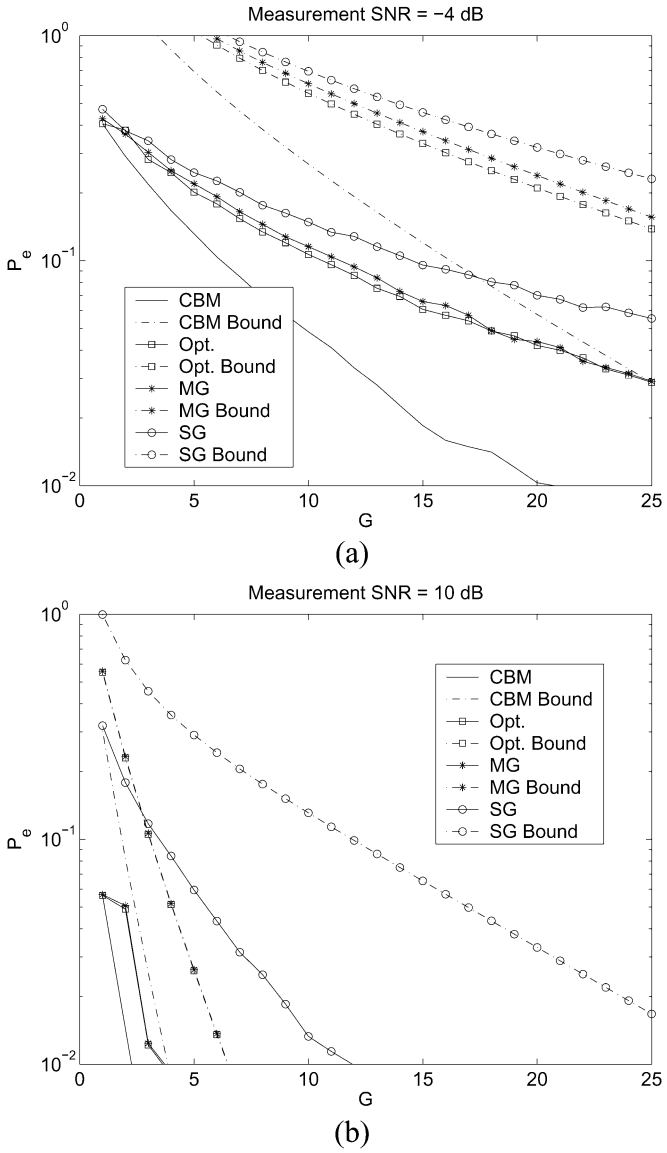


Fig. 7. Comparison of $P_{e,\mathrm{ideal}}$ from simulations with $P_{e,\mathrm{ideal}}$ Chernoff bounds for the various classifiers. (a) $\mathrm{SNR}_{\mathrm{meas}} = -4$ dB. (b) $\mathrm{SNR}_{\mathrm{meas}} = 10$ dB.

taken, which will limit the number of independent node measurements G. However, the effective G could be increased by processing multiple independent measurements over time.

The stationarity and Gaussianity assumptions on target signal statistics would likely be violated in practice. However, stationary modeling of signal *statistics* still allows for a fair amount of variability in target signature *realizations* across nodes. Appropriate models for the *dynamics* of target statistics could be combined with the techniques in this paper to develop adaptive distributed classifiers. Our numerical results indicate that even though the measured target signals were most likely non-Gaussian, second-order statistics are still adequate for reliable classification. Under the assumption of statistical independence across SCRs, non-Gaussian modeling of feature vectors in each SCR could be readily incorporated.

Finally, other approaches to classification, including tree-structured classifiers [10], subspace-based approaches

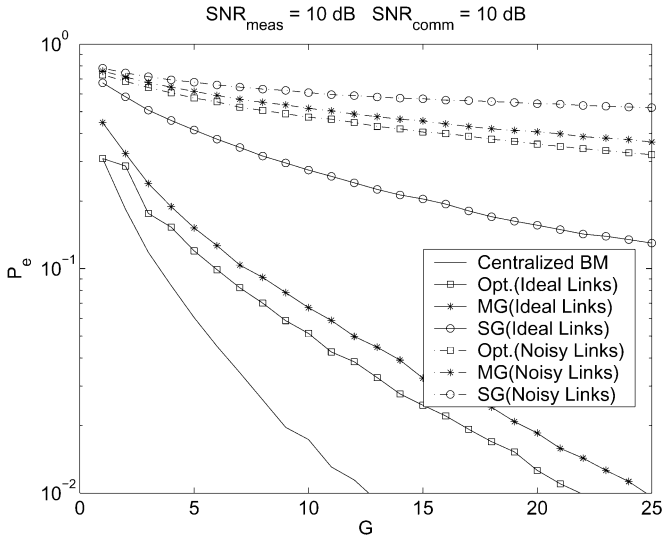


Fig. 8. $P_e(G)$ for $M = 3$ targets at $\mathrm{SNR}_{\mathrm{meas}} = 10$ dB and $\mathrm{SNR}_{\mathrm{comm}} = 10$ dB. The distributed classifiers under ideal and noisy communication links are compared with the centralized benchmark.

[11], [12], and suboptimal fusion schemes [13] could also be integrated into the proposed distributed architecture. The assumption of distinct targets could be relaxed by incorporating other attributes (such as orientation) in the classifiers, and the assumption of *a priori* knowledge of the number of targets could be relaxed by including it as a parameter in a composite hypothesis testing framework [8].

REFERENCES

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, vol. 4, 2001, pp. 2033–2036.

[2] S. Kumar, F. Zhao, and D. Shepherd, Eds., "Special issue on collaborative signal and information processing in microsensor networks," in *IEEE Signal Process. Mag.*, Mar. 2002, pp. 13–14.

[3] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification, tracking of targets in micro-sensor networks," in *IEEE Signal Process. Mag.*, Mar. 2002, pp. 17–29.

[4] A. D'Costa and A. M. Sayeed, "Collaborative signal processing for distributed classification in sensor networks," in *Lecture Notes in Computer Science* , F. Zhao and L. Guibas, Eds. Berlin, Germany, Apr. 2003, Proc. IPSN'03, pp. 193–208.

[5] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of Gaussian space–time sources in wireless sensor networks," *IEEE J. Sel, Areas. Commun., Special Issue on Sensor Networks*, vol. 22, no. 6, pp. 1026–1036, Aug. 2004.

[6] A. Sayeed, "A statistical signal modeling framework for wireless sensor networks," Univ. Wisconsin, Madison, WI, Tech. Rep. ECE-1-04, 2004.

[7] J. G. Proakis, *Digitial Communications*, 3rd ed. New York: McGraw-Hill, 1995.

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

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[10] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

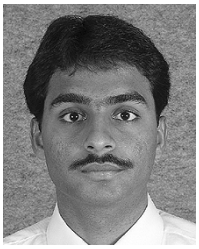
[11] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen–Loeve expansion to feature selection and ordering," *IEEE Trans. Comput.*, vol. C-19, pp. 311–318, Apr. 1970.

[12] S. Watanabe and N. Pakvasa, "Subspace method to pattern recognition," in *Proc. 1st Int. Conf. Pattern Recognit.*, Feb. 1973, pp. 25–32.

[13] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–238, Mar. 1998.

Jayesh H. Kotecha (M'04) received the B.E. degree in electronics and telecommunications from the College of Engineering, Pune, India, in 1995, and the M.S. and Ph.D. degrees from the State University of New York, Stony Brook, in 1996 and 2001, respectively.

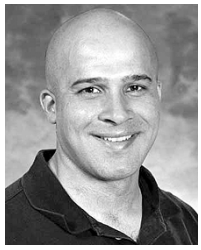
From January 2002 to August 2004, he was with the Wireless Communications Research Group, University of Wisconsin–Madison as a Postdoctoral Researcher. Currently, he is a Principal Staff Systems Engineer with Freescale Semiconductor, Inc. (formerly, Motorola Semiconductors), Austin, TX. His research interests are in the areas of wireless communications, sensor networks, statistical and adaptive signal processing, and particle filters.



Vinod Ramachandran received the B.Tech. degree in instrumentation engineering from Anna University, Chennai, India, in 2002, and the M.S. degree in electrical engineering from the University of Wisconsin–Madison in 2004.

He is currently working at Motorola India, Bangalore. His research interests include wireless communications, networking, and signal processing for sensor networks.

Mr. Ramachandran received the Donald and Esther Procknow Distinguished Graduate Fellowship from the University of Wisconsin in 2002.



Akbar M. Sayeed (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin–Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana–Champaign, in 1993 and 1996, respectively, all in electrical engineering.

From 1991 to 1996, he was a Research Assistant in the Coordinated Science Laboratory, University of Illinois. From 1996 to 1997, he was a Postdoctoral Fellow at Rice University, Houston, TX. Since August 1997, he has been with the University of Wisconsin– Madison, where he is currently Associate Professor of Electrical and Computer Engineering. His research interests are in wireless communications, statistical signal processing, information theory, and networks.

Dr. Sayeed received a Schlumberger Fellowship from 1992 to 1995 and the Robert T. Chien Memorial Award in 1996 for his doctoral research at the University of Illinois. He received the National Science Foundation (NSF) CAREER Award in 1999, the Office of Naval Research (ONR) Young Investigator Award in 2001, and a University of Wisconsin Grainger Junior Faculty Fellowship in 2003. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002.