

# Kernel-Based Learning of Decision Fusion in Wireless Sensor Networks

Gernot Fabeck, Rudolf Mathar  
Institute of Theoretical Information Technology  
RWTH Aachen University  
D-52056 Aachen, Germany  
Email: {fabeck,mathar}@ti.rwth-aachen.de

**Abstract**—The problem of decision fusion in wireless sensor networks for distributed detection applications has mainly been considered in scenarios where sensor observations are conditionally independent and both local sensor statistics as well as wireless channel conditions are available for fusion rule design. In this paper, kernel-based learning algorithms for the design of decision fusion rules are presented when no such prior knowledge is available. The fusion center receives a collection of labeled decision vectors from the sensor nodes and employs a discrete version of the method of kernel smoothing which exploits the ordinal nature of local sensor decisions. The aim is to arrive at fusion rules which are Bayes risk consistent, i.e., asymptotically optimal as the number of training samples tends to infinity. The kernel-based learning approach is applied to the problem of distributed detection of a deterministic signal in correlated Gaussian noise. Numerical results obtained by simulation show that the kernel-based fusion rules show good performance also for finite sample sizes.

**Keywords:** Decision fusion, wireless sensor networks, distributed detection, kernel-based learning.

## I. INTRODUCTION

One of the primary applications of wireless sensor networks is the detection of phenomena of interest in the monitored environment, e.g., absence or presence of a target [1], [2]. The wireless sensor nodes typically operate on limited energy budgets and are consequently subject to communication constraints, resulting in a finite number of bits each sensor node can transmit to the data sink before it runs out of power. In order to extend sensor network lifetime, preprocessing of measured raw data at the sensors and transmission of summary messages is recommended. In the parallel fusion topology, the sensor nodes process their observations independently and make preliminary decisions about the state of the observed environment. The sensors transmit the local decisions to a fusion center that combines the received messages and computes the final detection result. The problem of optimally designing the fusion rule according to the joint distribution of local sensor decisions as well as wireless channel conditions with respect to an overall performance criterion is called the problem of decision fusion.

Decision fusion for distributed detection is a well-developed field of research that traces back to the early work of Chair and Varshney [3]. Over the last two decades, the problem of decision fusion in sensor networks has been investigated

under a variety of different aspects like e.g. performance criteria, network topologies, and channel models [4]–[7]. In [8], the authors consider the design of optimal fusion rules in the practically important case that sensor decision rules are fixed. The derived fusion rules rely on the joint conditional probability density functions of all sensor observations as well as all local sensor decision rules and are valid for arbitrary network topologies and different performance criteria. In [9], Unnikrishnan and Veeravilli present a suboptimal solution to the fusion problem in the case of correlated observations under the assumption that the fusion center has access to partial statistical information about quantized observations in the form of lower order moments. Using deflection as a performance criterion, they obtain fusion rules which are optimal in the class of linear-quadratic detectors.

Despite the host of investigated scenarios, the majority of the literature relies on rather strong assumptions about the underlying statistical model. In the area of wireless sensor networks, these assumptions may become infeasible because one encounters random deployment of sensor nodes in complex and unamenable sensing environments. In such scenarios, previous knowledge of the underlying statistical model of sensor observations and wireless channel conditions may not be available for system design. To tackle these difficulties, adaptive approaches are recommended which facilitate the configuration of distributed detection systems on the spot.

The adaptive approaches presented in [10]–[13] are all based on estimating the local sensor error probabilities (i.e., the local probability of false alarm and the local probability of miss) in order to use these estimates in the Chair-Varshney fusion rule. The main limitations of these approaches are on the one hand the assumption of conditionally independent observations and on the other hand the restriction to hard decision quantization at the local detectors. Furthermore, error-free communication between the sensors and the fusion center is assumed.

An alternative framework for the design of decision fusion algorithms is provided by the field of supervised learning in sensor networks [14]. Supervised learning refers to learning from labeled samples of the underlying unknown probability distribution. In this paper, we present kernel-based learning algorithms for the design of fusion rules for distributed detection in the parallel fusion topology. After deployment of the network, the fusion center receives from the sensor nodes

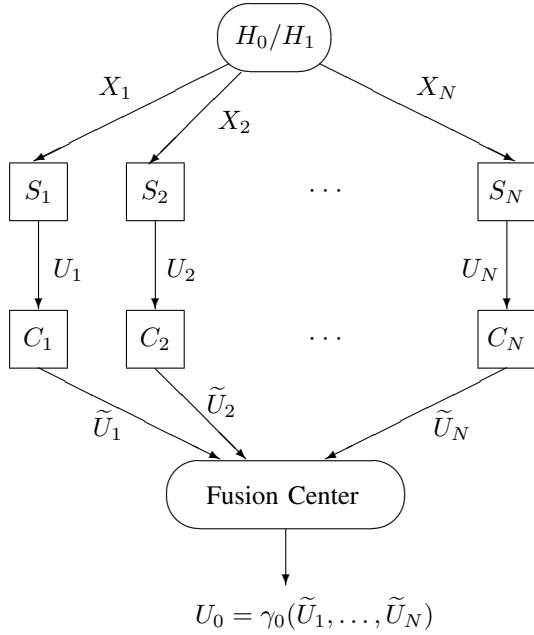


Fig. 1. Parallel fusion network with noisy channels.

a certain number of decision vectors which are labeled with the true underlying hypothesis by an external observer. For the subsequent learning phase we employ a discrete version of the method of kernel smoothing that takes advantage of the ordinal nature of local sensor decisions. Doing so, we obtain fusion rules which are Bayes risk consistent, i.e., asymptotically optimal for large training sets, but also show good performance for small sample sizes.

The remainder of this paper is organized as follows. In Section II, the problem of decision fusion for distributed detection in the parallel fusion network with noisy channels is stated. In Section III, we formulate kernel-based learning of decision fusion rules in the so-called sampling paradigm. In Section IV, we apply the kernel-based learning approach to the problem of distributed detection of a deterministic signal in correlated Gaussian noise and present numerical results obtained by simulation. Finally, we conclude in Section V.

## II. PARALLEL FUSION NETWORK WITH NOISY CHANNELS

The problem of decision fusion for distributed detection in the parallel fusion network with noisy channels can be stated as follows (see Fig. 1). We consider a binary hypothesis testing problem with hypotheses  $H_0$  and  $H_1$  indicating the state of the observed environment, e.g., absence or presence of a target. The actual state of the environment is described by a binary-valued random variable  $Y \in \{0, 1\}$  and associated prior probabilities

$$\begin{aligned} \pi_0 &= P(H_0) = P(Y = 0), \\ \pi_1 &= P(H_1) = P(Y = 1). \end{aligned} \quad (1)$$

In order to infer the true state of nature, a network of  $N$  distributed sensors  $S_1, \dots, S_N$  receive measurement data

$$\mathbf{X} = (X_1, \dots, X_N)' \in \mathcal{X}_1 \times \dots \times \mathcal{X}_N, \quad (2)$$

which are assumed to be generated according to either  $H_0$  or  $H_1$ , the two hypotheses under test. Due to the distributed nature of the problem, the sensors process their respective observations  $X_j$  independently by forming local decisions

$$U_j = \gamma_j(X_j), \quad j = 1, \dots, N. \quad (3)$$

In the general case of  $M$ -ary quantization, the local sensor decision rules  $\gamma_j$  are mappings

$$\gamma_j : \mathcal{X}_j \rightarrow \{1, \dots, M\}, \quad j = 1, \dots, N. \quad (4)$$

The structure of optimal sensor decision rules under a variety of conditions was investigated by Warren and Willett in [15]. It is important to note that under reasonable assumptions the local sensor decisions are of an ordinal nature, i.e., the larger the value of local decision  $U_j$  the more decides sensor  $S_j$  in favour of one of the two hypotheses.

Upon local detection, the sensor nodes transmit a vector of local decisions

$$\mathbf{U} = (U_1, \dots, U_N)' \in \{1, \dots, M\}^N \quad (5)$$

to the fusion center in order to perform decision combining. The communication channels between the wireless sensors and the fusion center may be subject to noise and interference. We follow an approach described by Cheng *et al.* [16] and model the communication link  $C_j$  between sensor  $S_j$  and the fusion center by a discrete noisy channel with transition matrix  $\mathbf{T}$ . We assume the channel transition matrix  $\mathbf{T} = (T_{kl})_{1 \leq k, l \leq M}$  to be a quadratic  $M \times M$  matrix with the  $kl$ th entry defined as

$$T_{kl} = P(\tilde{U}_j = k | U_j = l), \quad k, l \in \{1, \dots, M\}, \quad (6)$$

where  $\sum_{k=1}^M T_{kl} = 1$  for any  $l \in \{1, \dots, M\}$ . Due to the noisy channels, the fusion center receives a vector of potentially distorted decisions

$$\tilde{\mathbf{U}} = (\tilde{U}_1, \dots, \tilde{U}_N)' \in \{1, \dots, M\}^N. \quad (7)$$

The received decisions are combined to yield the final decision  $U_0 = \gamma_0(\tilde{U}_1, \dots, \tilde{U}_N)$ , where the fusion rule  $\gamma_0$  is a binary-valued mapping

$$\gamma_0 : \{1, \dots, M\}^N \rightarrow \{0, 1\}. \quad (8)$$

The sensor network detection performance is measured in terms of the global probability of error

$$\begin{aligned} P_e &= P(\gamma_0(\tilde{\mathbf{U}}) \neq Y) \\ &= \pi_0 P_f + \pi_1 P_m, \end{aligned} \quad (9)$$

which can be written as a weighted sum of the global probability of false alarm  $P_f = P(U_0 = 1 | H_0)$  and the corresponding global probability of miss  $P_m = P(U_0 = 0 | H_1)$ .

Since the decision fusion problem can be viewed as a hypothesis testing problem at the fusion center with received

local decisions being the observations, the Bayes optimal fusion rule  $\gamma_0^*$  takes the form

$$\gamma_0^*(\tilde{\mathbf{u}}) = \operatorname{argmax}_{k=0,1} P(H_k|\tilde{\mathbf{u}}) \quad (10)$$

$$= \operatorname{argmax}_{k=0,1} \pi_k \cdot p(\tilde{\mathbf{u}}|H_k), \quad (11)$$

where  $P(H_k|\tilde{\mathbf{u}})$  is the posterior probability of hypothesis  $H_k$  given the received decision vector  $\tilde{\mathbf{u}}$  and  $p(\tilde{\mathbf{u}}|H_k)$  is the joint probability mass function of the received decision vector under hypothesis  $H_k$ ,  $k = 0, 1$ . The minimum probability of error associated with the optimal fusion rule  $\gamma_0^*$  is given by the Bayes risk

$$P_e^* = P(\gamma_0^*(\tilde{\mathbf{U}}) \neq Y). \quad (12)$$

The implementation of the Bayes optimal fusion rule requires previous knowledge of either the posterior probabilities in (10) or of both the prior probabilities and the conditional joint probability mass functions in (11). Accordingly, the design of fusion rules based on a set of empirical samples can be done in two conceptually different ways. Whether the posterior probabilities are estimated directly or if estimates of the conditional joint probability mass functions are used via Bayes' rule, one talks about the diagnostic paradigm or the sampling paradigm, respectively [17].

In the following section, we present a kernel-based learning approach to the design of fusion rules in the sampling paradigm, i.e., the conditional joint probability mass functions of the received decision vectors under each hypothesis are estimated and plugged into (11).

### III. KERNEL-BASED LEARNING OF FUSION RULES

Approaches to estimate the joint probability mass function of an  $N$ -dimensional discrete random vector of  $M$ -ary data have to deal with the difficulty that the joint probability distribution is in general characterized by  $M^N - 1$  free parameters corresponding to the  $M^N$  different outcomes that can be observed. For large  $N$ , this results in a prohibitive amount of necessary training data for parameter estimation. In the decision fusion literature, this problem was circumvented for the hard decision case either by assuming complete knowledge of all correlations between the binary local decisions, or by allowing only the presence of restricted correlation structures that can be indexed by a single parameter [18], [19].

In order to cope with unknown and arbitrary correlation structures for  $M$ -ary local decisions, we use an extension of kernel smoothing for multivariate binary distributions which was introduced by Aitchison and Aitken [20]. Extending their approach to  $M$ -ary valued discrete random vectors allows the trained fusion rule to adapt to the correlation structure of the multivariate discrete distribution, paving the way for Bayes risk consistent fusion rules. Furthermore,  $M$ -ary quantization enables to take advantage of the above mentioned ordinary nature of local sensor decisions.

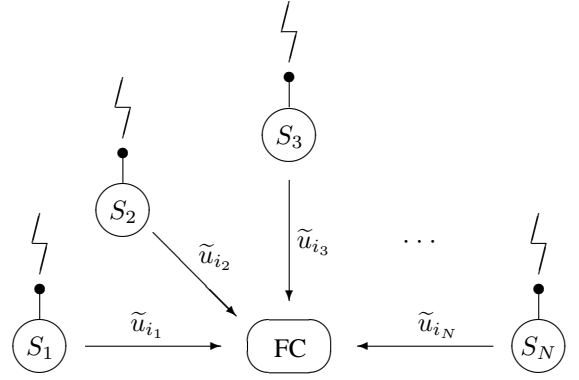


Fig. 2. Sensor nodes  $S_1, \dots, S_N$  transmitting the components of the decision vector  $\tilde{\mathbf{u}}_i = (\tilde{u}_{i1}, \dots, \tilde{u}_{iN})'$  to the fusion center (FC).

#### A. Kernel-based estimation

As an operational requirement for fusion rule design via supervised learning, the fusion center receives a certain number of decision vectors from the sensor nodes after deployment of the network (see Fig. 2). During this phase, each of the received decision vectors is assumed to be labeled with the true underlying hypothesis by an external observer. Formally, we assume that a training set

$$D_n = \{(\tilde{\mathbf{u}}_i, y_i)\}_{i=1}^n \subset \{1, \dots, M\}^N \times \{0, 1\} \quad (13)$$

of potentially distorted decision vectors  $\tilde{\mathbf{u}}_i \in \{1, \dots, M\}^N$  assigned with the label  $y_i \in \{0, 1\}$  of the true underlying hypothesis is available at the fusion center. The samples in the training set  $D_n$  are assumed to be drawn i.i.d. from the joint probability mass function  $p(\tilde{\mathbf{u}}, y)$  of the received decision vector  $\tilde{\mathbf{u}}$  and the underlying hypothesis  $y$ .

The fusion center uses the training set  $D_n$  in combination with a discrete kernel function

$$K : \{1, \dots, M\}^N \rightarrow \mathbb{R}_+ \quad (14)$$

to form estimates of the conditional joint probability mass functions  $p(\tilde{\mathbf{u}}|H_k)$  according to

$$\hat{p}(\tilde{\mathbf{u}}|H_0) = \hat{p}(\tilde{\mathbf{u}}|D_{n_0}, \boldsymbol{\lambda}_0) = \frac{1}{n_0} \sum_{D_{n_0}} K(\tilde{\mathbf{u}}|\tilde{\mathbf{u}}_i, \boldsymbol{\lambda}_0), \quad (15)$$

$$\hat{p}(\tilde{\mathbf{u}}|H_1) = \hat{p}(\tilde{\mathbf{u}}|D_{n_1}, \boldsymbol{\lambda}_1) = \frac{1}{n_1} \sum_{D_{n_1}} K(\tilde{\mathbf{u}}|\tilde{\mathbf{u}}_i, \boldsymbol{\lambda}_1), \quad (16)$$

where  $D_{n_k} \subset D_n$  is the subset of decision vectors with label  $k$ ,  $n_k$  is the number of decision vectors in the subset  $D_{n_k}$ , and  $\boldsymbol{\lambda}_k = (\lambda_{k1}, \dots, \lambda_{kN})'$  is a vector of smoothing parameters for  $k = 0, 1$ . The kernel-based estimates (15) and (16) are plugged into the Bayes optimal rule (11), yielding the fusion rule

$$\gamma_0(\tilde{\mathbf{u}}) = \operatorname{argmax}_{k=0,1} \hat{\pi}_k \cdot \hat{p}(\tilde{\mathbf{u}}|H_k). \quad (17)$$

The estimates  $\hat{\pi}_0$  and  $\hat{\pi}_1$  of the prior probabilities are chosen to be

$$\hat{\pi}_0 = \frac{n_0}{n}, \quad \hat{\pi}_1 = \frac{n_1}{n}. \quad (18)$$

For the discrete kernel function  $K$ , we employ a product kernel according to

$$K(\tilde{\mathbf{u}}|\tilde{\mathbf{u}}_i, \boldsymbol{\lambda}_k) = \prod_{j=1}^N K_j(\tilde{u}_j|\tilde{u}_{i,j}, \lambda_{k_j}) \quad (19)$$

where  $K_j(\tilde{u}_j|\tilde{u}_{i,j}, \lambda_{k_j})$  is the univariate kernel component for the  $j$ th component of  $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_N)'$  based on the  $j$ th component of the training vector  $\tilde{\mathbf{u}}_i = (\tilde{u}_{i1}, \dots, \tilde{u}_{iN})'$ . The simplest product kernel is composed from the nominal Aitchison-Aitken kernel components [20]

$$K(u|v, \lambda) = \begin{cases} \lambda & \text{if } u = v \\ (1 - \lambda)/(M - 1) & \text{if } u \neq v \end{cases} \quad (20)$$

However, the Aitchison-Aitken kernel does not allow use of the ordinal nature of local sensor decisions since it does not take the distance between the local decisions into account. A variety of univariate kernel functions which allow use of the ordinal scale of sensor decisions is given in Table I.

### B. Computing the smoothing parameters

Since our objective is to optimize sensor network detection performance, the choice of the smoothing parameters should be based on the aim of binary hypothesis testing. That means the smoothing parameters should be chosen according to a criterion which is based on the separation of the two hypotheses under test. We follow an approach due to Tutz and Groß [21] that is based on the notion of discriminant loss functions.

Let  $p(\tilde{\mathbf{u}}, y)$  denote the joint probability mass function of the received decision vector and the label of the underlying hypothesis and let  $\hat{p}(\tilde{\mathbf{u}}, y)$  denote its estimate. The family of functions measuring the discrimination loss associated with the estimate  $\hat{p}(\tilde{\mathbf{u}}, y)$  is given by

$$\mathcal{L}[p(\tilde{\mathbf{u}}, y), \hat{p}(\tilde{\mathbf{u}}, y)] = \sum_{\tilde{\mathbf{u}}} p(\tilde{\mathbf{u}}) \tilde{\mathcal{L}}[p(y|\tilde{\mathbf{u}}), \hat{p}(y|\tilde{\mathbf{u}})], \quad (21)$$

where  $p(\tilde{\mathbf{u}})$  denotes the marginal probability of  $\tilde{\mathbf{u}}$  and  $p(y|\tilde{\mathbf{u}})$ ,  $\hat{p}(y|\tilde{\mathbf{u}})$  denotes the posterior probability of  $y$  and the estimated posterior probability of  $y$ , respectively. The discriminant loss function (21) is a weighted sum of the conditional loss function  $\tilde{\mathcal{L}}$  which connects the loss to the hypothesis testing problem. The function  $\tilde{\mathcal{L}}$  penalizes the deviation of the estimated posterior probability  $\hat{p}(y|\tilde{\mathbf{u}})$  from the true posterior probability  $p(y|\tilde{\mathbf{u}})$ .

Examples of conditional loss functions are the Kullback-Leibler divergence

$$\tilde{\mathcal{L}}[p(y|\tilde{\mathbf{u}}), \hat{p}(y|\tilde{\mathbf{u}})] = \sum_y p(y|\tilde{\mathbf{u}}) \cdot \log \frac{p(y|\tilde{\mathbf{u}})}{\hat{p}(y|\tilde{\mathbf{u}})} \quad (22)$$

and the quadratic loss

$$\tilde{\mathcal{L}}[p(y|\tilde{\mathbf{u}}), \hat{p}(y|\tilde{\mathbf{u}})] = \sum_y (p(y|\tilde{\mathbf{u}}) - \hat{p}(y|\tilde{\mathbf{u}}))^2. \quad (23)$$

Direct evaluation of the discriminant loss function (21) requires knowledge of the true underlying distribution  $p(\tilde{\mathbf{u}}, y)$ . Since this knowledge is assumed to be not available, it is

replaced by the empirical knowledge of the training set  $D_n$ . Accordingly, the smoothing parameters are determined by minimizing the cross-validators estimate of the discrimination loss

$$\hat{\mathcal{L}}(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1) = \sum_{i=1}^n \mathcal{L}[\delta_{(\tilde{\mathbf{u}}_i, y_i)}, \hat{p}(\tilde{\mathbf{u}}, y|D_n \setminus \{(\tilde{\mathbf{u}}_i, y_i)\}, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1)], \quad (24)$$

where  $\delta_{(\tilde{\mathbf{u}}_i, y_i)}$  is the one-point measure putting the total mass at sample point  $(\tilde{\mathbf{u}}_i, y_i) \in D_n$  and  $D_n \setminus \{(\tilde{\mathbf{u}}_i, y_i)\}$  is the reduced training set where the observation  $(\tilde{\mathbf{u}}_i, y_i)$  is excluded. For the estimated joint probability mass function we employ

$$\hat{p}(\tilde{\mathbf{u}}, y|D_n, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1) = \begin{cases} \hat{\pi}_0 \cdot \hat{p}(\tilde{\mathbf{u}}|D_{n_0}, \boldsymbol{\lambda}_0) & \text{if } y = 0 \\ \hat{\pi}_1 \cdot \hat{p}(\tilde{\mathbf{u}}|D_{n_1}, \boldsymbol{\lambda}_1) & \text{if } y = 1 \end{cases} \quad (25)$$

Using the Kullback-Leibler divergence (22) or the quadratic loss (23) as conditional loss function, determination of the smoothing parameters according to

$$(\boldsymbol{\lambda}_0^*, \boldsymbol{\lambda}_1^*) = \underset{(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1)}{\operatorname{argmin}} \hat{\mathcal{L}}(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1) \quad (26)$$

results in a Bayes risk consistent fusion rule (17), i.e., its probability of error  $P_e$  converges with probability one to the probability of error  $P_e^*$  of the Bayes optimal fusion rule (12) as the size of the training set  $D_n$  tends to infinity [21].

### C. Example: Quadratic loss function

We give an example of the objective function (24) to be minimized for the special case of quadratic loss (23). Inserting expressions (25) in (24) and after some calculation, we obtain the objective function

$$\begin{aligned} \hat{\mathcal{L}}(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1) = & \sum_{D_{n_0}} \left[ \left( 1 - \frac{\hat{\pi}_0 \cdot \hat{p}(\tilde{\mathbf{u}}_i|D_{n_0} \setminus \{\tilde{\mathbf{u}}_i\}, \boldsymbol{\lambda}_0)}{\hat{p}(\tilde{\mathbf{u}}_i)} \right)^2 + \right. \\ & \left. + \left( \frac{\hat{\pi}_1 \cdot \hat{p}(\tilde{\mathbf{u}}_i|D_{n_1} \setminus \{\tilde{\mathbf{u}}_i\}, \boldsymbol{\lambda}_1)}{\hat{p}(\tilde{\mathbf{u}}_i)} \right)^2 \right] + \\ & \sum_{D_{n_1}} \left[ \left( 1 - \frac{\hat{\pi}_1 \cdot \hat{p}(\tilde{\mathbf{u}}_i|D_{n_1} \setminus \{\tilde{\mathbf{u}}_i\}, \boldsymbol{\lambda}_1)}{\hat{p}(\tilde{\mathbf{u}}_i)} \right)^2 + \right. \\ & \left. + \left( \frac{\hat{\pi}_0 \cdot \hat{p}(\tilde{\mathbf{u}}_i|D_{n_0} \setminus \{\tilde{\mathbf{u}}_i\}, \boldsymbol{\lambda}_0)}{\hat{p}(\tilde{\mathbf{u}}_i)} \right)^2 \right] \quad (27) \end{aligned}$$

which has to be minimized over the smoothing parameter vectors  $\boldsymbol{\lambda}_0$  and  $\boldsymbol{\lambda}_1$ . The expression  $\hat{p}(\tilde{\mathbf{u}}_i)$  denotes the relative frequency of the decision vector  $\tilde{\mathbf{u}}_i$  in the training set  $D_n$ .

## IV. NUMERICAL RESULTS

In this section, we provide numerical results by applying the presented kernel-based learning approach to the problem of distributed detection of a deterministic signal in correlated Gaussian noise. The effect of correlated Gaussian noise at the local sensors on the overall performance of the distributed detection system was studied by Aalo *et al.* [22]–[24]. As expected, they found that positive correlation between sensor observations tends to decrease overall system performance.

TABLE I  
ORDINAL UNIVARIATE KERNELS

Modified Habbema kernel	$K(u v, \lambda) = (1 - \lambda)^{ u-v ^\gamma}$	$\lambda \in [0, 1], \quad \gamma \in (0, \infty]$
Uniform kernel	$K(u v, \lambda) = \begin{cases} (1 - \lambda)/ T(v)  & \text{if } u \in T(v) \\ \lambda & \text{if } u = v \\ 0 & \text{else} \end{cases}$	$\lambda \in [0, 1]$
	where $T(v) = \{w \mid  v - w  \leq k, \quad v \neq w\}$	
Geometrical kernel	$K(u v, \lambda) = \begin{cases} \lambda & \text{if } u = v \\ \frac{1}{2}\lambda(1 - \lambda)^{ u-v } & \text{if } u \neq v \end{cases}$	$\lambda \in [0, 1]$
Aitken kernel	$K(u v, \lambda) = \begin{cases} \lambda & \text{if } u = v \\ (1 - \lambda)/2^{ u-v } & \text{if } u \neq v \end{cases}$	$\lambda \in [0, 1]$

### A. Joint distribution of sensor observations

We assume that the observations  $\mathbf{X} = (X_1, \dots, X_N)'$  at the local sensors of the distributed detection system are conditionally distributed according to

$$\begin{aligned} H_0 : \mathbf{X} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\ H_1 : \mathbf{X} &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \quad (28)$$

where  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the multivariate normal distribution with mean vector  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)'$  and covariance matrix  $\boldsymbol{\Sigma} = (\Sigma_{ij})_{1 \leq i, j \leq N}$ . The covariance matrix  $\boldsymbol{\Sigma}$  describes the correlated background noise and the mean vector  $\boldsymbol{\mu}$  indicates the deterministic signal component under hypothesis  $H_1$ . The hypotheses  $H_0$  and  $H_1$  are assumed to be equally likely to occur, i.e.,  $\pi_0 = \pi_1 = \frac{1}{2}$ .

At each sensor, the local observation signal-to-noise ratio (SNR)  $\text{SNR}_j$  is given in dB by

$$\text{SNR}_j = 10 \log_{10}\left(\frac{\mu_j^2}{\sigma_j^2}\right), \quad j = 1, \dots, N, \quad (29)$$

where  $\sigma_j^2 = \Sigma_{jj}$  is the local noise variance.

### B. Local sensor decision rules

We consider the case of quaternary sensors, i.e., the local sensor decision rules are mappings

$$\gamma_j : \mathcal{X}_j \rightarrow \{1, \dots, 4\}, \quad j = 1, \dots, N. \quad (30)$$

Motivated by the work of Chen and Papamarcou [25], we assume quantization of the local log-likelihood ratio

$$L_j = \log \frac{f_j(X_j|H_1)}{f_j(X_j|H_0)}, \quad j = 1, \dots, N, \quad (31)$$

where  $f_j(\cdot|H_k)$  denotes the marginal probability density function of observation  $X_j$  under hypothesis  $H_k$ ,  $k = 0, 1$ . The local log-likelihood ratios  $L_j$  are again Gaussian random

variables with conditional marginal distributions according to

$$\begin{aligned} H_0 : L_j &\sim \mathcal{N}\left(-\frac{\mu_j^2}{2\sigma_j^2}, \frac{\mu_j^2}{\sigma_j^2}\right), \\ H_1 : L_j &\sim \mathcal{N}\left(\frac{\mu_j^2}{2\sigma_j^2}, \frac{\mu_j^2}{\sigma_j^2}\right). \end{aligned} \quad (32)$$

The quantization of the local log-likelihood ratios  $L_j$  is done symmetrically resulting in local sensor decisions

$$U_j = \gamma_j(X_j) = \begin{cases} 1 & \text{if } -\infty < L_j \leq -\frac{\mu_j^2}{2\sigma_j^2} \\ 2 & \text{if } -\frac{\mu_j^2}{2\sigma_j^2} < L_j \leq 0 \\ 3 & \text{if } 0 < L_j \leq \frac{\mu_j^2}{2\sigma_j^2} \\ 4 & \text{if } \frac{\mu_j^2}{2\sigma_j^2} < L_j < \infty \end{cases}. \quad (33)$$

Given these fixed sensor decision rules, the objective is to implement a fusion rule with minimum probability of error. In the simulation, the fusion rule is constructed by a product of geometrical kernel functions (see Table I) and computation of the smoothing parameters is done using the quadratic loss function (27). For simplicity, we assume error-free communication channels.

We consider networks of  $N = 5, \dots, 50$  sensors collecting training sets of  $n = 50$  labeled decision vectors. Naturally, the resulting probability of error of the distributed detection system is a random variable because the fusion rule is trained on a set of random samples. So we calculate the mean probability of error by averaging over 100 independent simulation runs for each size of the network.

### C. Gaussian noise with geometrically decaying correlations

First, we consider Gaussian noise with geometrically decaying correlations, i.e., we assume that the correlation coefficients between  $X_i$  and  $X_j$  are given by

$$\rho_{ij} = \rho^{|i-j|}, \quad i, j = 1, \dots, N. \quad (34)$$

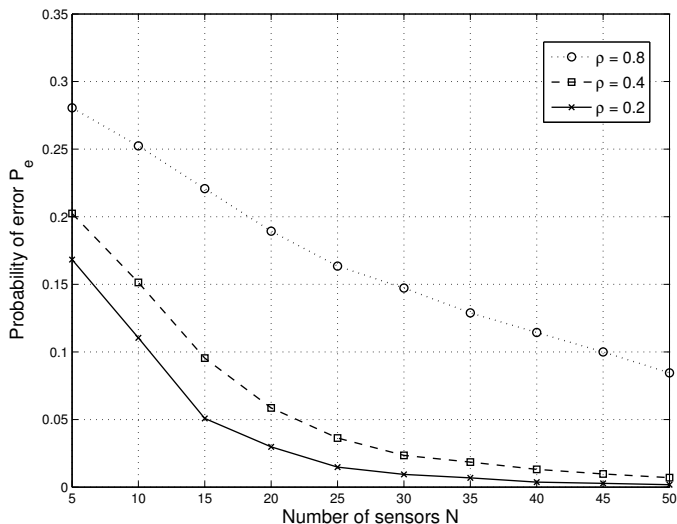


Fig. 3. Average probability of error of the kernel-based fusion rule for Gaussian noise with geometrically decaying correlations. The local observation SNR at each sensor is 2 dB.

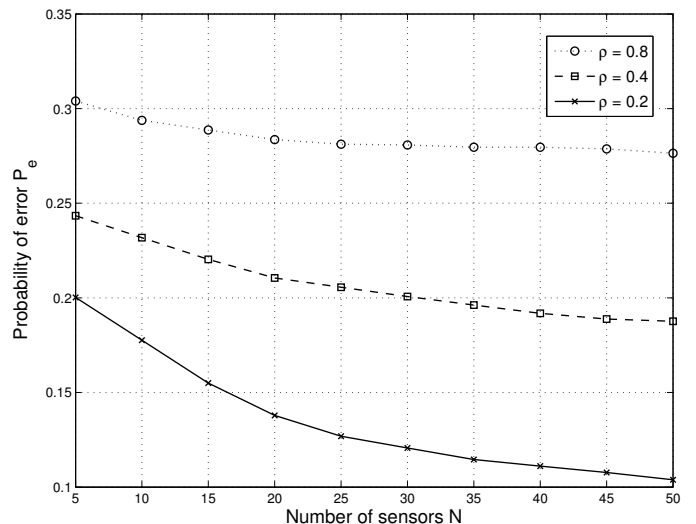


Fig. 4. Average probability of error of the kernel-based fusion rule for equicorrelated Gaussian noise. The local observation SNR at each sensor is 2 dB.

Assuming equal noise variance at the sensors, i.e.,  $\sigma_j^2 \equiv \sigma^2$ , the covariance matrix  $\Sigma$  has the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix}. \quad (35)$$

In this case, the degree of correlation is uniquely parameterized by  $\rho \in [0, 1]$ . This correlation model could be a reasonable approximation for some real-world situations.

Fig. 3 illustrates the average probability of error of the distributed detection system implementing the kernel-based fusion rule for various values of the correlation parameter  $\rho$ . It is obvious that with increasing  $\rho$  both system performance as well as the benefit of additional sensors decreases.

#### D. Equicorrelated Gaussian noise

Next, we consider equicorrelated Gaussian noise, i.e., we consider a covariance matrix  $\Sigma$  of the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}. \quad (36)$$

Again, the degree of correlation is parameterized by  $\rho \in [0, 1]$ .

The numerical results depicted in Fig. 4 illustrate once again the deteriorating effect of strong positive correlations on sensor network detection performance. However, the kernel-based learning approach to fusion rule design provides robust results in the strong correlation case also.

#### E. Gaussian noise with random correlations

In order to illustrate the adaptability of the kernel-based learning approach, we finally consider the case of Gaussian

noise where the degree of correlation between the observations is random. In this case, the covariance matrix  $\Sigma$  is a random matrix where the entries  $\Sigma_{ij}$  are random variables. There are several methods for generating random covariance matrices with prespecified properties. We follow a recent approach developed in [26] which is easy to implement.

First, we consider a scenario where the non-diagonal entries of  $\Sigma$  are random but the diagonal entries are fixed. In particular, we assume  $\Sigma_{11} = \Sigma_{22} = \dots = \Sigma_{NN} = \sigma^2$ , resulting in fixed and identical local observation SNR at the sensors. Fig. 5 depicts the simulation results. The average probability of error of the kernel-based fusion rule approximately decays exponentially with the number of sensors  $N$ .

Finally, we allow the local observation SNR to be random, too. Consequently, it may differ from sensor to sensor. The distribution of the diagonal entries  $\Sigma_{11}, \dots, \Sigma_{NN}$  is chosen in a way that the local observation SNR's are randomly distributed between 0 and 5 dB. In correspondence with the results of Fig. 5, Fig. 6 shows an approximately exponential decay of the probability of error.

## V. CONCLUSIONS

In this paper, we have presented a kernel-based learning approach to the design of decision fusion rules for distributed detection in sensor networks. Using discrete kernel functions which exploit the ordinal nature of local sensor decisions and computing appropriate smoothing parameters, this approach resulted in Bayes risk consistent fusion rules which provide optimal detection performance in the asymptotic case. Numerical results obtained by application of the kernel-based learning approach to the problem of distributed detection of a deterministic signal in correlated Gaussian noise showed good performance of the obtained fusion rules also for moderately sized training sets.

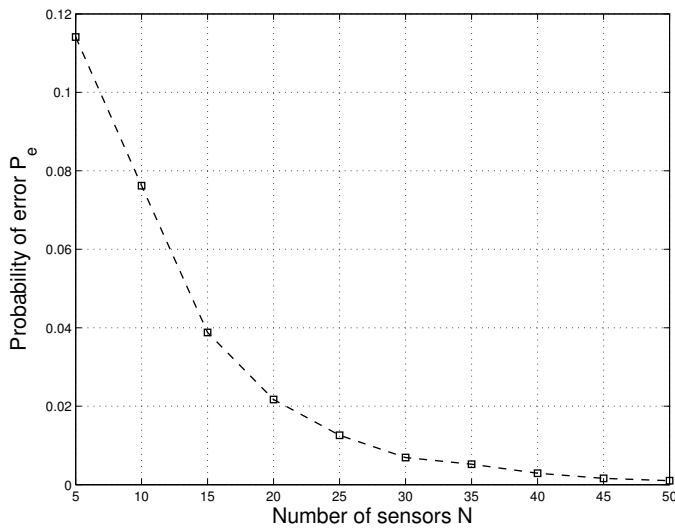


Fig. 5. Average probability of error of the kernel-based fusion rule for Gaussian noise with random correlations. The local observation SNR at each sensor is 2 dB.

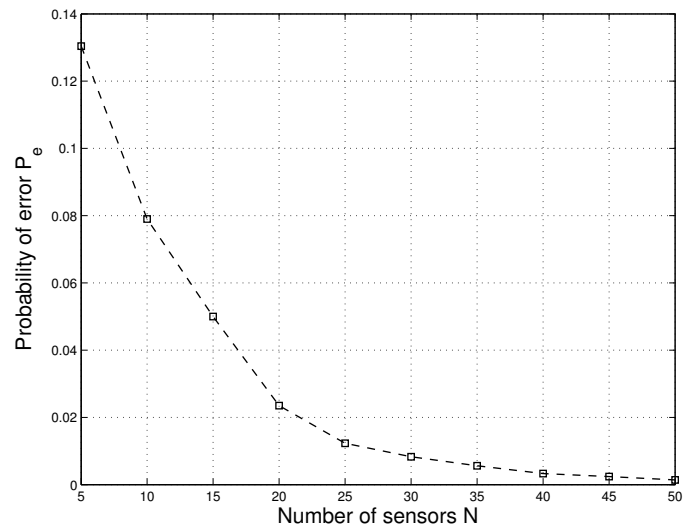


Fig. 6. Average probability of error of the kernel-based fusion rule for Gaussian noise with random correlations and random observation SNR. The SNR at each sensor randomly takes values between 0 and 5 dB.

## REFERENCES

- [1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets in distributed sensor networks," in *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17-29, March 2002.
- [2] J.-F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," in *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16-25, May 2007.
- [3] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 1, pp. 98-101, January 1986.
- [4] S. C. A. Thomopoulos, R. Viswanathan, and D. C. Bougoulas, "Optimal decision fusion in multiple sensor systems," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 23, no. 5, pp. 644-653, September 1987.
- [5] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 366-376, July 1988.
- [6] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," in *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3454-3458, December 2004.
- [7] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," in *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1018-1027, March 2006.
- [8] Y. Zhu and X. R. Li, "Optimal decision fusion given sensor rules," in *Proc. 1999 International Conference on Information Fusion*, Silicon Valley, CA, pp. 85-92, July 1999.
- [9] J. Unnikrishnan and V. V. Veeravalli, "Decentralized detection with correlated observations," in *Proc. 2007 Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 381-385, November 2007.
- [10] A. Naim and M. Kam, "On-line estimation of probabilities for distributed Bayesian detection," in *Automatica*, vol. 30, no. 4, pp. 633-642, 1994.
- [11] N. Ansari, E. S. H. Hou, B. Zhu, and J. Chen, "Adaptive fusion by reinforcement learning for distributed detection systems," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 524-531, April 1996.
- [12] X. Zhu, M. Kam, and Q. Zhu, "Adaptive Bayesian decision fusion," in *Proc. 1997 IEEE Conference on Decision and Control*, San Diego, CA, pp. 5004-5009, December 1997.
- [13] G. Mirjalily, Z. Luo, T. N. Davidson, and É. Bossé, "Blind adaptive decision fusion for distributed detection," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 1, pp. 34-52, January 2003.
- [14] S. N. Simic, "A learning theory approach to sensor networks," in *IEEE Pervasive Computing*, vol. 2, no. 4, pp. 44-49, 2003.
- [15] D. Warren and P. Willett, "Optimum quantization for detector fusion: Some proofs, examples, and pathologies," *J. Franklin Inst.*, vol. 336, pp. 323-359, 1999.
- [16] Q. Cheng, B. Chen, and P. K. Varshney, "Detection performance limits for distributed sensor networks in the presence of nonideal channels," in *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 1554-1563, October 2006.
- [17] A. P. Dawid, "Properties of diagnostic data distributions," in *Biometrics*, vol. 32, no. 3, pp. 647-658, September 1976.
- [18] E. Drakopoulos and C. C. Lee, "Optimum multisensor fusion of correlated local decisions," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 593-606, July 1991.
- [19] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," in *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 98-101, January 1986.
- [20] J. Aitchison and C. G. Aitken, "Multivariate binary discrimination by the kernel method," in *Biometrika*, vol. 63, no. 3, pp. 413-420, 1976.
- [21] G. Tutz and H. Groß, "Discrete kernels, loss functions and parametric models in discrete discrimination: A comparative study," in *Mathematical Methods of Operations Research*, vol. 42, pp. 217-230, 1995.
- [22] V. Aalo, R. Viswanathan, and S. C. A. Thomopoulos, "A study of distributed detection with correlated sensor noise," in *Proc. 1987 IEEE Global Telecommunications Conference*, Tokyo, Japan, pp. 259-262, November 1987.
- [23] V. Aalo and R. Viswanathan, "On distributed detection with correlated sensors: Two examples," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 414-421, May 1989.
- [24] V. Aalo and R. Viswanathan, "Asymptotic performance of a distributed detection system in correlated Gaussian noise," in *IEEE Transactions on Signal Processing*, vol. 40, no. 1, pp. 211-213, January 1992.
- [25] P. Chen and A. Papamarcou, "Likelihood ratio partitions for distributed signal detection in correlated Gaussian noise," in *Proc. 1987 IEEE International Symposium on Information Theory*, Whistler, Canada, p. 118, September 1995.
- [26] H. Joe, "Generating random correlation matrices based on partial correlations," in *Journal of Multivariate Analysis*, vol. 97, pp. 2177-2189, 2006.