Formation Detection with Wireless Sensor Networks

I. Ch. Paschalidis, Boston University
Wuyang Dai, Boston University
Dong Guo, Boston University

We consider the problem of detecting the formation of a set of wireless sensor nodes based on the pairwise measurements of signal strength corresponding to all transmitter/receiver pairs. We assume that formations take values in a discrete set and develop a composite hypothesis testing approach which uses a Generalized Likelihood Test (GLT) as the decision rule. The GLT distinguishes between a set of probability density function (pdf) families constructed using a custom pdf interpolation technique. The GLT is compared with the simple Likelihood Test (LT). We also adapt one prevalent supervised learning approach, Multiple Support Vector Machines (MSVMs), and compare it with our probabilistic methods. Due to the highly variant measurements from the wireless sensor nodes, and these methods' different adaptability to multiple observations, our analysis and experimental results suggest that GLT is more accurate and suitable for formation detection. The formation detection problem has interesting applications in posture detection with Wireless Body Area Networks (WBANs), which is extremely useful in health monitoring and rehabilitation. Another valuable application we explore concerns autonomous robot systems.

Categories and Subject Descriptors: I.5.2 [**Pattern Recognition**]: Design Methodology

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Wireless sensor networks, formation detection, wireless body area networks, autonomous robot systems.

ACM Reference Format:
Paschalidis, I. Ch., Dai, W., and Guo, D. 2014. Formation Detection with Wireless Sensor Networks. ACM Trans. Sensor Netw. 1, 1, Article 1 (January 2014), 18 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

Wireless Body Area Networks (WBANs) consist of small, battery powered, wireless sensor nodes attached (or implanted) to the human body. Interesting sensors include pacemakers, implantable cardioverter-defibrillators (ICDs), pulse oximeters, glucose level sensors, sensors to monitor the heart (ECG, blood pressure, etc.), thermometers, or sensors to monitor body posture. Several devices incorporating such sensing capabilities and wireless connectivity exist today (at least as prototypes) [Otto et al. 2006;

Research partially supported by the NSF under grants EFRI-0735974, CNS-1239021, IIS-1237022, by the DOE under grant DE-FG52-06NA27490, by the ARO under grants W911NF-11-1-0227 and 61789-MA-MUR, and by the ONR under grant N00014-10-1-0952.
Authors' addresses: I. Ch. Paschalidis, Department of Electrical and Computer Engineering and the Division of Systems Engineering, Boston University, 8 Saint Mary's Street, Boston, MA 02215, USA, e-mail: yannisp@bu.edu, url: http://ionia.bu.edu/; W. Dai, Department of Electrical and Computer Engineering, Boston University, 8 Saint Mary's Street, Boston, MA 02215, USA, email: wydai@bu.edu; D. Guo, Center for Information and Systems Engineering, 15 Saint Mary's Street, Brookline, MA 02446, USA, email: dong.dongguo@gmail.com.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

Latre et al. 2007]. The emergence of WBANs could potentially revolutionize health monitoring and health care, allowing for instance remote uninterrupted monitoring of patients in their normal living environment. Broad applications are envisioned in medicine, military, security, and the workplace [Jovanov et al. 2005; Torfs et al. 2007].

One particular service that is of great interest in many of these application contexts is to detect body posture. We give a few examples. Work related back injuries are a frequent source of litigation. Many such injuries can be avoided if an incorrect posture for lifting heavy packages can be detected early on and corrected by proper training. In another example, the inhabitants of a smart home will be able to control different functions (such as heating or cooling, lighting, etc.) merely by gesture. Such functionality can be important for the elderly and the disabled. An additional application involves monitoring inhabitants of assisted living facilities; body posture reveals a lot about the "state" of an individual and could alert caregivers in case of emergency (e.g., body falling to or lying on the stairs [Lai et al. 2010]).

The basic idea is to place WBAN devices on different parts of the body, say, wrist, ankle, shoulder, knee, etc., and to detect the posture of the body through the *formation* of the wireless sensor nodes, i.e., the relative positions of these nodes. The premise of this paper is that the formation of the wireless sensor nodes is reflected by the *Received Signal Strength Indicators (RSSI)* at WBAN nodes. As our experiments show, this relationship indeed holds, but in a rather complicated way. In particular, the RSSI signatures of a formation (or posture) do not have a simple dependence on the pairwise distances of the WBAN nodes [Ray et al. 2006; Paschalidis and Guo 2009]. Instead, they are correlated among different WBAN pairs, do not follow a standard noise model, and they also depend on the time and the location of the body and other subtle aspects (e.g., the thickness of clothes). This is the reason we focus on measurement-based methods, including probabilistic classifiers and supervised learning approaches.

The problem at hand has a wider applicability than the WBAN setting. The techniques we develop are also applicable in detecting (and controlling) the formation of robot swarms deployed in the interior of a building. In indoor deployments, the mapping from RSSI to distance is erratic and unpredictable, requiring the more sophisticated classification or hypothesis testing techniques we develop (see also [Ray et al. 2006; Paschalidis and Guo 2009; Li et al. 2012]).

In this work, we assume that formations take values in a discrete set. We develop a new method that constructs a pdf family for each formation by leveraging a generalized joint pdf interpolation scheme; a simpler interpolation scheme was proposed by [Bursal 1996]. We then formulate the formation detection problem as the composite hypothesis testing problem that determines the most likely pdf family from which RSSI measurements originate. To solve this problem we propose a *Generalized Likelihood Test (GLT)*. We compare this approach to two alternatives. The first is a simple *Likelihood Test (LT)* applied to a standard hypothesis testing problem which associates a single pdf (rather than a pdf family) to each formation. LT is also widely applied to pattern detection in spatial data [Kulldorff 1997; 2001; Neill 2012]. The second alternative is a supervised learning approach – the *Multiple Support Vector Machine (MSVM)* [Cortes and Vapnik 1995; Duan and Keerthi 2005].

Comparisons between these three methods are carried out by simulations and also by experiments involving actual sensor nodes. We conduct the testing for both a single observation (RSSI measurements at a certain time) and multiple observations (a sequence of measurements). The results show that our pdf family construction coupled with GLT has several potential advantages compared to the two alternatives:

— it results in better handling of multiple observations;
— is robust to measurement uncertainty;

— and is computationally more efficient for multi-class classification when the number of classes is large.

Measurement uncertainty, in particular, is due to both changes in the environment where measurements are taken, and, most importantly, due to systematic changes in the measurement process, e.g., misalignment of the sensors between the training and the formation detection phases.

The rest of the paper is organized as follows. In Section 2, we survey recent research on posture detection with WBANs and on formation detection for robot swarms. We also discuss several applications of GLT and SVM-based methods. In Section 3, we formulate our problem and introduce our notation. In Section 4, we introduce the two decision rules for the hypothesis testing formulation. Section 5 describes the MSVM approach. In Section 6 and Section 7, we describe simulation and experimental results and compare the various methods. We end with some concluding remarks in Section 8.

Notation: We use bold lower case letters for vectors and bold upper case letters for matrices. All vectors are column vectors and we write $\mathbf{x} = (x_1, \cdots, x_n)$ for economy of space. Transpose is denoted by prime.

2. RELATED WORK

Recent developments in sensor technology make wearable sensors and the resulting body area network applicable to a variety of scenarios. Body posture detection in particular, has been studied for different purposes and with different approaches. One application concerns detecting a fall of the monitored individual, which is useful in protecting senior citizens [Lai et al. 2010]. Although video monitoring or alarms with buttons could offer alternatives, they have their own limitations. The former one raises privacy concerns while the latter one requires the senior person's consciousness after falling. A WBAN solution, however, does not suffer from these limitations [Lai et al. 2010].

[Farella et al. 2006] constructs a custom-designed WBAN for posture detection. The constructed WBAN uses accelerometers to acquire information about the posture, and then makes a classification according to a pre-set table. The accelerometers are also used in [Foerster et al. 1999] for posture detection in ambulatory monitoring. As shown in these previous works, it is common to use accelerometers as the main source of relevant data ([Lai et al. 2010; Farella et al. 2006; Foerster et al. 1999]). Accelerometers indeed provide accurate measurements for quick movements and thus are more suitable for motion detection. However, for posture detection, they need an inference step to "derive" posture from motion, which makes the detection more complicated and highly dependent on the logical rules implemented in this inference step. In Section 7, we conduct experiments to demonstrate the insufficiency of accelerometers in certain situations. As we will see, RSSI provides complementary information which can be leveraged to render posture detection more accurate and robust to measurement noise.

The work of [Quwaider and Biswas 2008] provides a novel approach for posture detection by using the relative proximity information (based on RSSI) between sensor nodes. Compared to accelerometer-based methods, this approach is not limited to activity intensive postures, such as walking and running, but also works for low activity postures such as sitting and standing. This is one of the main reasons we elect to use RSSI signals for posture detection. [Quwaider and Biswas 2008] uses a technique involving a Hidden Markov Model (HMM). Our approach is quite different as it does not exploit temporal relations among measurements and this contributes to increased robustness to measurement noise. Our rationale is that capturing temporal relations among measurements necessitates a more sophisticated model, which in turn requires more samples for training the model parameters and an increased computational ef-

fort. At the same time, a sophisticated model has more system parameters, which elevates the risk of having modeling errors and increases the sensitivity to measurement noise, either due to systematic (e.g., sensor misalignment) or random causes.

In addition to the posture detection, detecting the formation of sensor nodes in a wireless sensor network has found a major application in robot swarms. Recent developments in swarm robotics introduce new applications such as navigation and terrain exploration by self-assembled systems [Batalin and Sukhatme 2002; Christensen et al. 2007; Nouyan et al. 2008]. Self-assembled systems are attractive as they can more easily adapt to various unforeseen changes to their environment. For example, consider a robot swarm navigating an unknown area. The robots have to deploy themselves and assume multiple formations to fit the various parts of the terrain they are exploring [Christensen et al. 2007]. To make adjustments, knowing their current formation is inevitably needed and thus accurate formation detection becomes essential.

We next turn to surveying work related to the methodologies we employ. The *Support Vector Machine (SVM)* is a binary classifier well known for its generally good performance in many applications [Burges 1998]. To adapt SVM from binary classification to multiclass classification, it is common to apply SVM between each pair of classes and then employ a simple majority vote to make the final decision [Duan and Keerthi 2005]. We call this extension *Multiple SVM (MSVM)*. In addition to machine learning, maximum likelihood-based techniques are also used for classification. In the related but simpler problem of sensor location detection, a simple hypothesis testing approach was introduced in [Ray et al. 2006]. For the same application a more robust approach involving composite hypothesis testing was developed in [Paschalidis and Guo 2009]. Yet a different approach was introduced in [Li et al. 2012].

The present work essentially extends the line of work in [Ray et al. 2006; Paschalidis and Guo 2009; Li et al. 2012] to the more complex problem of formation detection. The key salient difference of our present work from the localization work in [Ray et al. 2006; Paschalidis and Guo 2009; Li et al. 2012] is that localization utilizes the (marginal) pdf of measurements from a single sensor at a set of receivers whereas formation detection needs the *joint* pdf of measurements from multiple sensors at a single receiver. As we will see, this requires several innovations including appropriate procedures for pdf estimation and interpolation. Further, our earlier work focused on establishing GLT optimality under certain conditions and optimally placing the multiple receivers whereas in the present paper the focus is on the formation detection application and the pros and cons of alternative sensor modalities and classification approaches.

3. FORMULATION

Consider k sensors, where one of them is the receiver and the rest are transmitters, and let $\mathscr{C} = \{1, \ldots, C\}$ be a discrete set of their possible formations. In practice, the positions of the sensors take values in a continuous space and one can argue that formations are also continuous. However, for many applications, including the ones discussed in the Introduction, we are interested in distinguishing between relatively few formations which characterize the "state" of the underlying physical system (e.g., the body, the robot swarm).

The discretization of formations is in line with our earlier sensor localization work [Ray et al. 2006; Paschalidis and Guo 2009; Li et al. 2012]. It makes the detection/classification problem more tractable but introduces the requirement that the techniques to be used should be robust enough and tolerant to mild or moderate perturbations. As mentioned in the Introduction, such perturbations cause systematic differences between measurements taken during the training and detection phases. To accommodate these differences, we take every element of $\mathscr{C}$ to represent a "family"

of similar looking formations that can be generated from a nominal formation subject to perturbations.

The RSSI measurements at the receiver are denoted by a column vector $\mathbf{y} \in \mathbb{R}^d$, where $d = k - 1$. In each of the methods we will present, the formation classifier is computed from a *training* set of RSSI measurements, and then we examine experimentally how well the classifiers generalize to additional measurements. Two types of methods for building the classifier will be considered next: a probabilistic hypothesis testing approach and MSVM.

4. PROBABILISTIC APPROACH

In the probabilistic approach, we treat each formation as a composite hypothesis associated with a family of pdfs in the space of the joint RSSI measurements. We use a family of pdfs for each formation in order to improve system robustness (e.g., with respect to time and location). The pdfs are first estimated from the training data, employing a technique combining a Parzen windowing scheme and Gaussianization [Erdogmus et al. 2004]. The pdf families are formed using a pdf interpolation technique that we have generalized from [Bursal 1996]. Finally, decisions are made according to the well-known (generalized) maximum likelihood test.

4.1. Multivariate density estimation

Suppose among the M samples, $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are associated with one formation in $\mathscr{C}$. Let $\mathbf{Y} = [\mathbf{y}_1 \mathbf{y}_2 \cdots \mathbf{y}_m]$. We view the measurements $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ as realizations of a random variable $\mathbf{y} = (y_1, y_2, \ldots, y_d)$. We first estimate the marginal pdfs of $\mathbf{y}$ denoted by $p_i(y_i)$, $i = 1, \ldots, d$, using Parzen windows (see Appendix A). The benefit of using Parzen windows is that the resulting $p_i(y_i)$'s are smoothed.

We then estimate the joint pdf using the Gaussianization method of [Erdogmus et al. 2004], the basic assumption (or approximation) of which is : when we transform the marginal distributions separately into Gaussian distributions, the joint distribution also becomes Gaussian. Specifically, we construct an element-wise Gaussianization function $\mathbf{h}(\mathbf{y}) = (h_1(y_1), h_2(y_2), \ldots, h_d(y_d))$, such that the marginal distributions of $\mathbf{z} = \mathbf{h}(\mathbf{y})$ are zero-mean Gaussian distributions. Then, we assume $\mathbf{z}$ is also jointly Gaussian, thus, its pdf can be determined from the sample covariance matrix $\mathbf{\Sigma_z}$. The joint pdf of $\mathbf{y}$ can therefore be estimated as ([Erdogmus et al. 2004]):

$$p(\mathbf{y}) = \frac{g_{\mathbf{\Sigma_z}}(\mathbf{h}(\mathbf{y}))}{|\nabla \mathbf{h}^{-1}(\mathbf{h}(\mathbf{y}))|} = g_{\mathbf{\Sigma_z}}(\mathbf{h}(\mathbf{y})) \prod_{i=1}^{d} \frac{p_i(y_i)}{g_1(h_i(y_i))}, \tag{1}$$

where $g_{\mathbf{\Sigma_z}}$ denotes a zero-mean multivariate Gaussian density function with covariance $\mathbf{\Sigma_z}$, p_i is the i-th marginal distribution of $\mathbf{y}$, and g_1 denotes a zero-mean univariate Gaussian density function with unit variance.

4.2. Interpolation of probability density functions

In order to construct a family of pdfs for each formation, we introduce an interpolation technique for probability density functions.

Let each $f_i(\mathbf{y})$, $i = 1, \ldots, N$, be a d-dimensional pdf with mean $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{K}_i$. Note that these are generally non-Gaussian pdfs. We call what follows the *linear interpolation* of these pdfs with a weight vector $\boldsymbol{\alpha}$, where the elements of $\boldsymbol{\alpha}$ are nonnegative and sum to one.

It is desirable that the mean and covariance of the interpolated pdf equal

$$\boldsymbol{\mu} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\mu}_i, \qquad \mathbf{K} = \sum_{i=1}^{N} \alpha_i \mathbf{K}_i. \tag{2}$$

Define a coordinate transformation for each $i = 1, \ldots, N$, so that given $\mathbf{y}$ (the target position, at which we are trying to evaluate the density), $\mathbf{y}_i$ is defined by

$$\mathbf{K}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i), \tag{3}$$

where $\mathbf{K}^{1/2}(\mathbf{K}^{1/2})' = \mathbf{K}$. The Jacobian of each transformation is expressed as

$$J_i = \sqrt{\det(\mathbf{K}_i \mathbf{K}^{-1})}. \tag{4}$$

The interpolation formula is then

$$f_{\boldsymbol{\alpha}}(\mathbf{y}) = \sum_{i=1}^{N} \alpha_i J_i f_i(\mathbf{y}_i). \tag{5}$$

This interpolation not only achieves property (2), but also preserves the "shape" information of the original pdfs to a large extent. For example, if the original pdfs are Gaussian, then the interpolated pdf is also Gaussian. This cannot be achieved by, say, a simple weighted sum of the original pdfs. The formula above was first given in [Bursal 1996], but formally only for cases satisfying $d = N$. We verify that the general case is also true in Appendix B.

It worth noting that given distinct models, one can devise several (and perhaps more sophisticated) alternatives to linear interpolation. Added sophistication, however, can substantially increase the computational overhead. Given that, as we will see, the linear interpolation yields pretty good experimental results, we elected to not explore alternative interpolation techniques.

4.3. LT and GLT

We associate a hypothesis H_j to each formation $j \in \mathscr{C}$. For each formation j, we collect measurements from different deployments of the nodes according to j in different environments (e.g., rooms of a building). The idea is to capture a variety of "modes" of the environment that can affect RSSI, as well as, sample a set of potential perturbations of sensor positions corresponding to a particular formation. For each set of measurements, we construct a pdf $f(\mathbf{y}|H_j)$ as outlined in Section 4.1. As in Section 4.2, we interpolate the pdfs corresponding to different deployments of formation j to end up with a pdf family $f_{\boldsymbol{\alpha}}(\mathbf{y}|H_j)$ characterizing this formation. As explained earlier, the key motivation for constructing pdf families is to gain in robustness with respect to small perturbations that would naturally arise in any deployment of a formation.

The maximum likelihood test (LT) is based on just a single pdf $f(\mathbf{y}|H_j)$ characterizing formation j. Using n observations (sets of RSSI measurements) $\mathbf{y}_1, \ldots, \mathbf{y}_n$, it identifies formation H_L if

$$L = \arg\max_{j \in \mathscr{C}} \prod_{i=1}^{n} f(\mathbf{y}_i|H_j). \tag{6}$$

The test we propose is a composite hypothesis test using the pdf families $f_{\boldsymbol{\alpha}}(\mathbf{y}|H_j)$. It uses the generalized likelihood test (GLT) which was shown to have desirable optimality properties in [Paschalidis and Guo 2009]. Specifically, it identifies formation H_L if

$$L = \arg\max_{j \in \mathscr{C}} \max_{\boldsymbol{\alpha}} \prod_{i=1}^{n} f_{\boldsymbol{\alpha}}(\mathbf{y}_i|H_j). \tag{7}$$

5. MULTIPLE SUPPORT VECTOR MACHINE

In this section we describe a classification approach using a *Support Vector Machine (SVM)*. An SVM is an excellent two-category classifier [Cortes and Vapnik 1995]. We

work with one pair of formations, l_1 and l_2, at a time. To find the support vectors, we solve the following dual form of the soft margin problem (see [Cortes and Vapnik 1995]):

$$\begin{aligned} \max \ & -\frac{1}{2}\sum_{i=1}^{M_1}\sum_{j=1}^{M_1}\alpha_i\alpha_j I_i I_j K(\mathbf{y}_i,\mathbf{y}_j)+\sum_{i=1}^{M_1}\alpha_i, \\ \text{s.t.} \ & \sum_{i=1}^{M_1}\alpha_i I_i = 0, \\ & 0 \le \alpha_i \le \Lambda, \end{aligned} \tag{8}$$

where $\mathbf{y}_i$'s are the original measurements, Λ is the penalty coefficient, $K(\cdot,\cdot)$ is the kernel function, $I_i = \pm 1$ is the label of sample i with 1 meaning formation l_1 and -1 meaning formation l_2, and M_1 is the total number of samples associated with either formation. Given a measurement $\mathbf{y}$, the SVM categorizes it by computing

$$I_{l_1 l_2}(\mathbf{y}) = \text{sign}(\sum_{i=1}^{M_1} I_i \alpha_i K(\mathbf{y},\mathbf{y}_i)), \tag{9}$$

where $I_{l_1 l_2}(\mathbf{y})$ denotes the output label. Again, 1 means formation l_1 and -1 means formation l_2.

We tried several commonly used kernel functions and ended up using the Gaussian radial basis function:

$$K(\mathbf{y}_1,\mathbf{y}_2) = \exp(-\frac{\|\mathbf{y}_1-\mathbf{y}_2\|^2}{2\sigma^2}). \tag{10}$$

For a C-class SVM, as in our case, we can apply $C(C-1)/2$ pairwise SVMs, and use a majority vote to make the final decision [Duan and Keerthi 2005]:

$$L = \arg\max_{i\in\mathscr{C}} \sum_{j\neq i} I_{ij}(\mathbf{y}). \tag{11}$$

Formula (11) is for a single observation classification. With multiple observations, we need another level of majority voting over n observations.

In summary, MSVM needs to run SVM several times to classify a given piece of test data and each run involves more than one class of training data. On the other hand for GLT (or LT), the calculation of the likelihood of test data for a certain hypothesis only needs the training data of that class. Given C classes, GLT needs C sets of models, one for each class. Each set is the outcome of interpolations. Suppose we discretize the possible values of $\boldsymbol{\alpha}$ in (7) and assume up to Γ possible values (including the originally estimated pdfs which correspond to $\boldsymbol{\alpha}$ being equal to a unit vector). Then, GLT needs $O(C\Gamma)$ amount of work to make a decision for each test input. On the other hand, MSVM performs $\binom{C}{2}$ binary classifications, which is on the order of $O(C^2)$. If we consider Γ as a constant with regard to C, the complexity of GLT grows linearly in C and we have the potential of requiring much less computational resources as C increases. In our experiments, however, due to the limitation of time and computational resources, we only performed tests involving a 3-class problem. For such a small C, our implementation of GLT took more time (but not substantially more) than SVM. To better understand at which values of C the above asymptotic analysis takes hold, we conducted a simulation experiment where samples of each class were drawn from a Gaussian distribution. We compared the running time of GLT and MSVM under different values of C. The experimental results support the analysis, however, the value of C

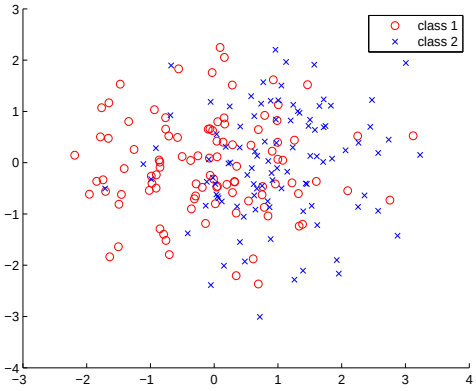

Fig. 1. Samples of points drawn from the two Gaussian distributions.

has to be large (on the order of 200) for GLT to take less time than MSVM. For smaller C's (below 200), GLT and MSVM are quite similar in terms of running time, with the former maintaining its other advantages over MSVM (e.g., robustness to measurement uncertainty).

6. SIMULATIONS

6.1. Setup

We first compare the methods we discussed using simulated data. We generate points in $\mathbb{R}^2$ drawn from two Gaussian distributions. Points of *class 1* are drawn from a Gaussian with mean $(0, 0)$ and covariance equal to the identity matrix. Points of *class 2* are drawn from a Gaussian with mean $(1, 0)$ and covariance equal to the identity matrix. Sample points drawn from these two distributions are shown in Figure 1.

We generated 100 training data points and 500 test data points per class. The LT and SVM algorithms were directly applied to these data. For GLT, the training data were randomly split into two subsets, each with an equal number of points. For each class we derived an empirical pdf of point locations within each one of the two subsets. We then applied the approach of Section 4.2 and constructed a pdf family for each class as the interpolation of the two empirical pdfs corresponding to the two subsets. The GLT was applied using these two (class 1 and class 2) pdf families. The whole process was repeated 100 times in order to eliminate any variability due to the randomly generated data. Figure 2 plots the average (over the 100 repetitions) classification accuracies achieved by the three algorithms as a function of the number of observations used. By classification accuracy, we denote the fraction of correctly classified test data in the test data set. The results show that even though all three methods perform equally well when a single observation is used, with multiple observations probabilistic methods (LT, GLT) achieve higher classification accuracies than SVM. GLT and LT perform similarly because samples of each class are drawn from a single pdf and there is no systematic difference between samples. Our next setup is aimed at highlighting the differences between GLT and LT.

In the above setting, the means of the two classes are fixed. We set up another simulation experiment with "uncertain" means, reflecting systematic differences between samples (e.g., due to sensor misalignment). More specifically, noise is added into one dimension of each mean vector as follows: the mean of class 1 is set to $(x_1, 0)$ where x_1 is uniformly distributed in the interval $[-5, 0]$, while the mean of class 2 is set to

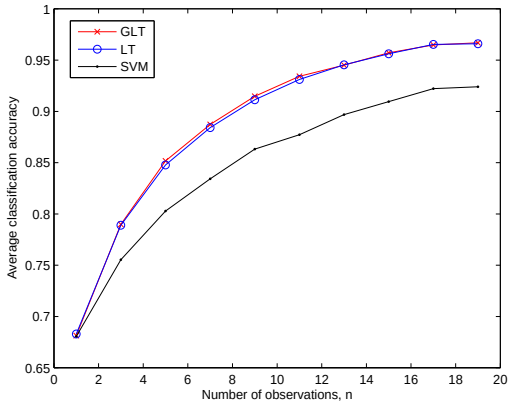

Fig. 2. Average classification accuracies of different methods on simulated data.

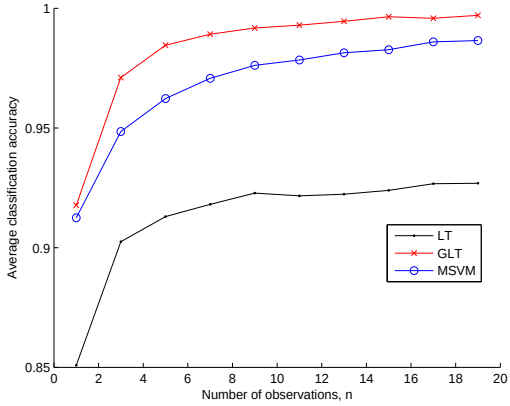

Fig. 3. Average classification accuracies of different methods on simulated data with uncertain means.

$(1, y_2)$ with y_2 uniformly distributed in the interval $[0, 5]$. Two training data sets are generated under the extreme values of the means, while the test data are generated under random mean values. We train the three classifiers as described earlier. The GLT classifier uses a pdf family for each class derived as the interpolation of the empirical pdfs built from each of the two training sets. The rationale for using the extreme values for the means during training is that, in practice, we ought to gather several sets of data (much more than just two) and data from the extreme distributions are likely to be among them. For this experiment, we plot the average classification accuracies achieved by the three algorithms in Figure 3. The results essentially validate our premise that led us to develop the GLT approach. They indicate that GLT is indeed more "robust" to systematic uncertainty than LT and SVM is substantially inferior to GLT.

6.2. Discussion

Our simulation results show that GLT and LT perform better for multi-observation test/classification. The intuition behind this is that in expressions (6) and (7), the like-

lihoods of different observations are multiplied together so that one large likelihood (corresponding to high confidence) can dominate others. Ideally, if the empirical pdf in LT (pdf family in GLT) is indeed the underlying density and the multiple observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are i.i.d., the multiplication $f(\mathbf{y}_1) \cdots f(\mathbf{y}_n)$ (correspondingly $f_{\boldsymbol{\alpha}}(\mathbf{y}_1) \cdots f_{\boldsymbol{\alpha}}(\mathbf{y}_n)$ for GLT) yields the joint density evaluated at the n observations. As a result, (6) becomes the likelihood ratio test using the joint distribution which is guaranteed to be optimal. Similarly, (7) becomes the GLT using the joint distribution and it also optimal under certain conditions [Paschalidis and Guo 2009]. On the other hand, in the MSVM approach, each observation (independent of our confidence level) simply adds one vote to a class.

In the simulation experiment with uncertain mean values, GLT outperforms LT because GLT has the ability to appropriately shift the density to fit the test data by making use of the free parameter in the pdf family constructed by interpolating the two (extremal) empirical pdfs. This is a unique characteristic of GLT and it results in GLT's robustness with respect to system parameters (i.e., the mean values in this case). It is worth noting that this characteristic of GLT requires the availability of the "extremal" distributions in the training data (or more loosely sufficient diversity in the training data) and is also affected by the mechanism of pdf interpolation.

7. EXPERIMENTS

7.1. Hardware

For our experiments we used the Intel Imote2 motes from Memsic Corp. to measure the RSSI. The Imote2 (2400-2483.5 MHz band) uses the Chipcon CC2420, IEEE 802.15.4 compliant, ZigBee-ready radio frequency transceiver integrated with an PXA271 micro-controller. Its radio can be tuned within the IEEE 802.15.4 channels, numbered from 11 (2.405 GHz) to 26 (2.480 GHz), each separated by 5 MHz. The RF transmission power is programmable from 0 dBm (1 mW) to -25 dBm. In order to reduce the signal variation for each posture, we tuned the RF transmission power to -25 dBm at channel 11.

In addition to RSSI, we also measure the angle formed by the trunk of a body and the ground using an Imote2 ITS400 sensor board which has an onboard 3-axis accelerometer. For this measurement, we only need 1-axis information.

7.2. Setups

Setups 1 and 2 target body posture while Setup 3 concerns robot swarm formation.

Setup 1. We use 4 sensors (measuring only RSSI) attached to the right upper chest, outside of left wrist, left pocket and left ankle. It is easy and convenient to attach sensors at these 4 body areas. Among all these sensors, the right upper chest one is used as the receiver while the rest are transmitters.

If the postures are very different (such as standing vs. bending forward), classification is much easier and all methods (GLT, LT and MSVM) show very high accuracy. To discern differences among the three methods we use the following three patterns which are not quite easy to differentiate:

— standing straight with hands aligned with the body (military standing at attention);
— standing straight with the two hands loosely held together in front of the body;
— and standing straight with the two arms folded in front of the chest.

It is quite obvious that with only accelerometers, these three similar postures are not separable.

We capture three sets of data from the sensors at different times. Each set contains all postures. Each of the three sets has roughly 1000 observations per posture. In each

experiment, we randomly select 200 samples per posture from these three data sets and samples from two of them are used for training, while samples from the remaining data set are used for testing. We repeat the experiment 60 times and report the classification accuracies in Figure 4.

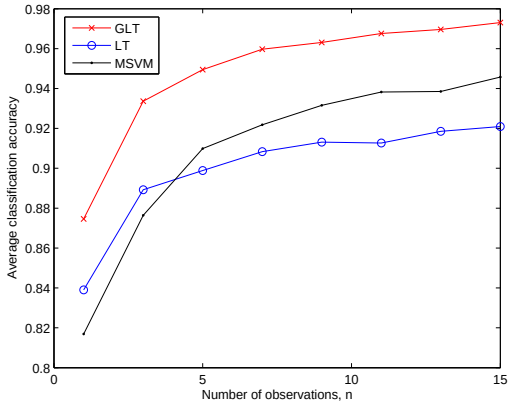

Fig. 4. Average classification accuracies of different methods on real sensor data under Setup 1.

The experimental results lead to the same conclusions as in Section 6. The advantage of GLT is obvious and as more observations are used the classification accuracy approaches 97%. We note that with GLT, it is possible to distinguish between very closely related postures which can broaden the applicability of posture detection.

Setup 2. In this setting, the previous four sensors are still used and attached to the same positions. In addition, one more sensor, which measures the inclination of the trunk of the body relative to the horizontal, is attached to the chest. We design three postures that require both the RSSI information and the angle information to differentiate them:

— standing straight with hands aligned with the body (military standing at attention);
— bending forward to almost 90 degrees;
— and lying down flat.

In view of the body formations, the first posture is exactly the same as the third one. However, they may imply a very different condition in a application where an elderly resident is monitored in her home. In particular, lying on the floor may be due to unconsciousness and is reason to alert emergency services.

The data collection here is similar to Setup 1. We collect three sets of data with 1000 observations per posture in each of the three data sets. From these, 200 samples per posture are randomly selected each time for performing the experiment. We plot average classification accuracies when using only accelerometer data in Figure 5 and when using both RSSI and accelerometer data in Figure 6.

It can be seen that using accelerometer data only does not lead to high classification accuracies. Yet, GLT outperforms LT and MSVM for $n \geq 5$ in those cases. By adding RSSI measurements we can achieve classification accuracies on the order of 91%–94% and differentiate postures quite well. It can also be seen that GLT performs better than the other two methods for smaller n and equally better with LT than MSVM for larger n.

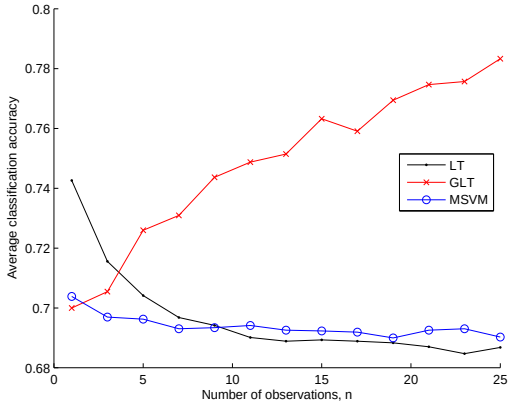

Fig. 5. Average classification accuracies of different methods with only accelerometer measurements under Setup 2.

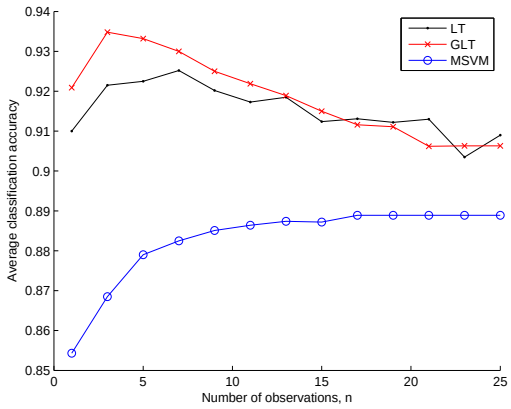

Fig. 6. Average classification accuracies of different methods with both RSSI and accelerometer measurements under Setup 2.

Setup 3. In this setting, we target a very different application: formation detection applied to robot swarms. We consider a swarm of robots roaming within a building and seek to detect the formation they are in, out of a discrete repertoire of possible formations. In our experimental setup we simply place sensors on the floor of the building according to three different formations: rectangle, parallelogram, and linear, which are shown in Figures 7 – 9. In these figures, Sensor* is the receiver while all other sensors are transmitters.

The rectangular and linear formations have been considered elsewhere in the literature and are suitable for a number of different applications [Christensen et al. 2007]. The parallelogram formation can be thought of being a "transition" formation between the rectangular and the linear. In this setup as well, our data collection procedure is exactly the same as the one described under Setup 1. We plot results from the three algorithms in Figure 10. The GLT method is again demonstrating consistently better accuracies than both LT and MSVM.

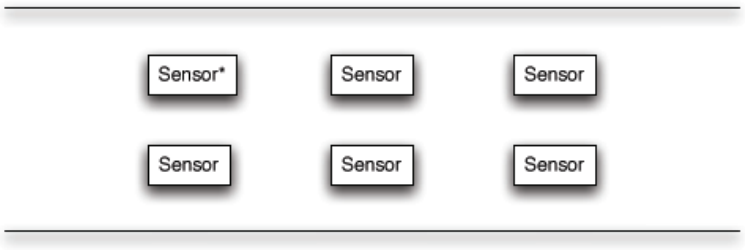

Fig. 7. Rectangle formation of robot swarm.

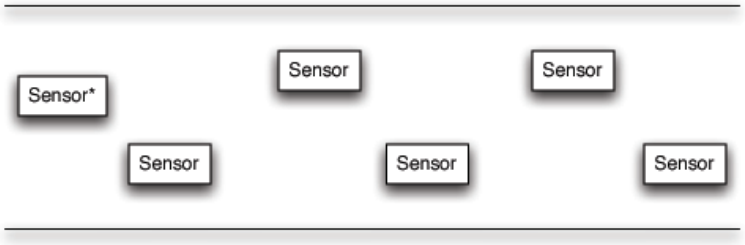

Fig. 8. Parallelogram formation of a robot swarm.

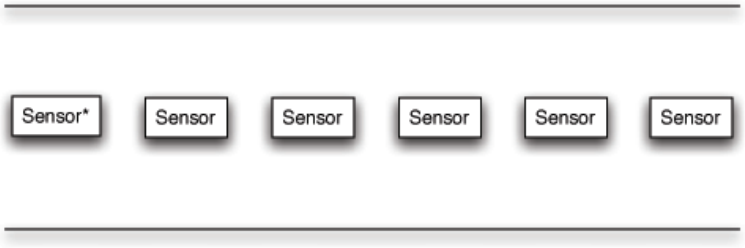

Fig. 9. Linear formation of a robot swarm.

While performing the various computations we observed that LT and GLT can be vulnerable to numerical precision errors. Specifically, when likelihoods of certain observations are small and we multiply many of them together, the result may end up being zero if sufficient precision accuracy is not used.

To further support the comparison between GLT and MSVM, we provide another plot showing the stability of the ranking of GLT over MSVM. Figure 11 shows the percentage of random tests for which GLT performs at least as well as MSVM. It can be seen that GLT has at least a 91.5% chance of performing equally well or better than MSVM. The results were derived for the robot swarm application. Similar observations hold for other experiments (as in Figure 2-6 and for n greater than 5).

All of these establish the usefulness of applying GLT on a broad range of formation detection applications. In our experimental results, the superior performance of GLT is due to its ability to better handle multiple observations and systematic uncertainty. For different scenarios, the main reason could be different. For example, results in Figure 5 support the claim that GLT handles better multiple observations. On the other hand, results at $n = 1$ in Figures 4 and 10, show that GLT produces better

predictions than LT and MSVM even when using a single sample; this is likely due to its tolerance to systematic uncertainty.

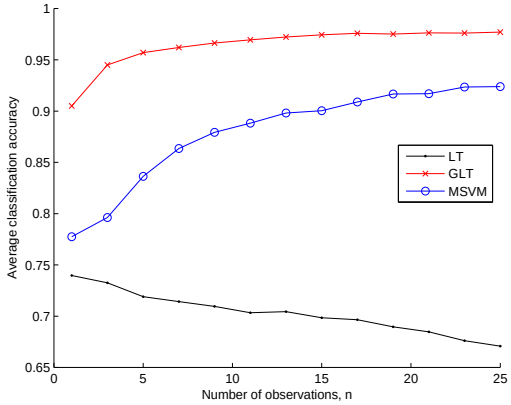

Fig. 10. Average classification accuracies of different formations of robot swarms under Setup 3.

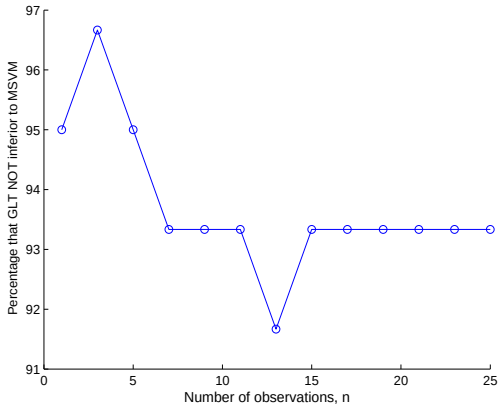

Fig. 11. Percentage of random tests that GLT performs at least as well as MSVM.

8. CONCLUSIONS

We considered the problem of formation detection with wireless sensor networks. This problem has various applications in human body posture detection and robot swarm formation detection. By using RSSI measurements between wireless devices, the problem is formulated as a multiple-pattern classification problem. We developed a probabilistic (hypothesis testing based) approach, the core of which includes the construction of a pdf family representation of formation features. We further analyzed and compared this algorithm (GLT) with LT and MSVM. The simulation and experimental results support the claim that GLT works better due to its ability to handle multiple observations and its robustness to systematic uncertainty.

The GLT approach can also be useful in detecting novel (i.e., not previously seen) postures. To that end, one can introduce a threshold on the value of the likelihood and declare a new posture if that value drops below the threshold. Such a test has similar optimality guarantees as GLT (see the analysis of movement detection in [Li et al. 2012]). MSVM on the other hand, partitions the feature space into subregions and is not suited to detecting novel postures.

APPENDIX

A. PARZEN WINDOW PDF ESTIMATION

We applied Parzen windows to estimate the marginal pdfs of our observation data. For a set of scalar samples $\{x_1, \cdots, x_N\}$ the Parzen windows estimate for the marginal pdf is

$$\hat{f}(x) = \frac{1}{N}\sum_{j=1}^{N} K_\sigma(x - x_j), \tag{12}$$

where the kernel function $K_\sigma(\cdot)$ is a Gaussian pdf with zero-mean and variance σ^2. The parameter σ controls the width of the kernel and is known as the kernel size. We use the default σ value that is optimal for estimating normal densities [Bowman and Azzalini 1997], which is

$$\sigma_{opt} = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma}n^{-\frac{1}{5}}$$

where $\hat{\sigma}$ is the standard deviations of the samples. This is a common and effective way of selecting the kernel size. There are also many other methods for bandwidth selection and a brief survey is included in [Jones et al. 1996].

B. VERIFICATION OF LINEAR INTERPOLATION

We verify that $f_{\boldsymbol{\alpha}}(\mathbf{y})$ (cf. (5)) is a pdf with mean $\boldsymbol{\mu}$ and variance $\mathbf{K}$. First, we verify that $f_{\boldsymbol{\alpha}}(\mathbf{y})$ is a probability measure:

$$\begin{aligned}
&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\boldsymbol{\alpha}}(\mathbf{y})dy_1\cdots dy_d \\
=&\sum_{i=1}^{N}\alpha_i\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} J_i f_i(\mathbf{y}_i)dy_1\cdots dy_d \\
=&\sum_{i=1}^{N}\alpha_i\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_i(\mathbf{y}_i)dy_{i1}\cdots dy_{id} \\
=&\sum_{i=1}^{N}\alpha_i = 1.
\end{aligned}$$

The first equality above is obtained by directly plugging in $f_{\boldsymbol{\alpha}}(\mathbf{y})$. The second above equality uses that J_i is the Jacobian of each transformation and $J_i dy_1\cdots dy_d = dy_{i1}\cdots dy_{id}$.

Then we verify the mean:

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathbf{y}f_{\boldsymbol{\alpha}}(\mathbf{y})dy_1\cdots dy_d$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} \alpha_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{y} J_i f_i(\mathbf{y}_i) dy_1 \cdots dy_d \\
&= \sum_{i=1}^{N} \alpha_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\mu}) f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} \\
&= \sum_{i=1}^{N} \alpha_i (\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{y}_i - \boldsymbol{\mu}_i) f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} \\
&\quad + \boldsymbol{\mu} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id}) \\
&= \boldsymbol{\mu} \sum_{i=1}^{N} \alpha_i = \boldsymbol{\mu}.
\end{aligned}
$$

The second equality above is due to (3). The fourth equality is due to $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{y}_i - \boldsymbol{\mu}_i) f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} = 0$, ($\boldsymbol{\mu}_i$ is the mean of $\mathbf{y}_i$) and $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} = 1$ ($f_i(\mathbf{y}_i)$ is a probability measure).

Lastly, we check the covariance matrix:

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{y}\mathbf{y}' f_{\boldsymbol{\alpha}}(\mathbf{y}) dy_1 \cdots dy_d \\
&= \sum_{i=1}^{N} \alpha_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{y}\mathbf{y}' J_i f_i(\mathbf{y}_i) dy_1 \cdots dy_d \\
&= \sum_{i=1}^{N} \alpha_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\mu})(\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\mu})' f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} \\
&= \sum_{i=1}^{N} \alpha_i (\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)' f_i(\mathbf{y}_i)(\mathbf{K}_i^{-1/2})'(\mathbf{K}^{1/2})' dy_{i1} \cdots dy_{id} \\
&\quad + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i)\boldsymbol{\mu}' f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} \\
&\quad + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{\mu}(\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i))' f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id} \\
&\quad + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{\mu}\boldsymbol{\mu}' f_i(\mathbf{y}_i) dy_{i1} \cdots dy_{id})
\end{aligned}
$$

$$= \sum_{i=1}^{N} \alpha_i (\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}\mathbf{K}_i(\mathbf{K}_i^{-1/2})'(\mathbf{K}^{1/2})' + \boldsymbol{\mu}\boldsymbol{\mu}')$$

$$= (\boldsymbol{\mu}\boldsymbol{\mu}' + \mathbf{K}) \sum_{i=1}^{N} \alpha_i = \boldsymbol{\mu}\boldsymbol{\mu}' + \mathbf{K}.$$

The second equality above is obtained by substituting $(\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\mu})$ for $\mathbf{y}$, which is derived from (3). The third equality is obtained by expanding all the terms in brackets and results in four terms. The first term simply calculates the variance of $\mathbf{y}_i$ and then scales it by some constants. This term turns out to be $\mathbf{K}^{1/2}\mathbf{K}_i^{-1/2}\mathbf{K}_i(\mathbf{K}_i^{-1/2})'(\mathbf{K}^{1/2})'$. The second and the third term turn to be zero because $\boldsymbol{\mu}_i$ is the mean of $\mathbf{y}_i$. The fourth term is constant.

Our verification is complete.

ACKNOWLEDGMENTS

The authors would like to thank Saikat Ray for useful discussions and Binbin Li and Yingwei Lin for helping with some of the experimental results. We also want to thank the anonymous referees whose comments and suggestions have helped us improve the paper.

REFERENCES

BATALIN, M. A. AND SUKHATME, G. S. 2002. Spreading out: A local approach to multi-robot coverage. In *Proceedings of the 6th International Symposium on Distributed Autonomous Robotics Systems*. Fukuoka, Japan.

BOWMAN, A. W. AND AZZALINI, A. 1997. *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, New York.

BURGES, C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2,* 2, 121–167.

BURSAL, F. H. 1996. On interpolating between probability distributions. *Applied Mathematics and Computation 77*, 213–244.

CHRISTENSEN, A. L., O'GRADY, R., AND DORIGO, M. 2007. Morphology control in a multirobot system. *IEEE Robotics & Automation Magazine 14*, 18–25.

CORTES, C. AND VAPNIK, V. 1995. Support-vector networks. *Machine learning 20,* 3, 273–297.

DUAN, K.-B. AND KEERTHI, S. S. 2005. Which is the best multiclass SVM method? an empirical study. *Multiple Classifier Systems*, 278–285.

ERDOGMUS, D., JENSSEN, R., RAO, Y., AND PRINCIPE, J. 2004. Multivariate density estimation with optimal marginal Parzen density estimation and Gaussianization. In *2004 IEEE Workshop on Machine Learning for Signal Processing*. 73–82.

FARELLA, E., PIERACCI, A., BENINI, L., AND ACQUAVIVA, A. 2006. A wireless body area sensor network for posture detection. In *Proceedings of the 11th IEEE Symposium on Computers and Communications*. Washington, DC, USA, 454–459.

FOERSTER, F., SMEJA, M., AND FAHRENBERG, J. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior 15,* 5, 571–583.

JONES, M., MARRON, J., AND SHEATHER, S. J. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American statistical association 91,* 433, 401–407.

JOVANOV, E., MILENKOVIC, A., OTTO, C., AND DE GROEN, P. 2005. A wireless body network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of Neuroengineering and Rehabilitation 2,* 6.

KULLDORFF, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods 26,* 6, 1481–1496.

KULLDORFF, M. 2001. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A 164,* 1, 61–72.

LAI, C., HUANG, Y., CHAO, H., AND PARK, J. 2010. Adaptive body posture analysis using collaborative multi-sensors for elderly falling detection. *IEEE Intelligent Systems*.

LATRE, B., BRAEM, B., MOERMAN, I., BLONDIA, C., REUSENS, E., JOSEPH, W., AND DEMEESTER, P. 2007. A low-delay protocol for multihop wireless body area networks. In *Fourth Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. Philadelphia, Pennsylvania.

LI, K., GUO, D., LIN, Y., AND PASCHALIDIS, I. C. 2012. Position and movement detection of wireless sensor network devices relative to a landmark graph. *IEEE Transactions on Mobile Computing 11,* 12, 1970–1982.

NEILL, D. B. 2012. Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society: Series B 74,* 2, 337–360.

NOUYAN, S., CAMPO, A., AND DORIGO, M. 2008. Path formation in a robot swarm: Self-organized strategies to find your way home. *Swarm Intell 2*, 1–23.

OTTO, C., MILENKOVIĆ, A., SANDERS, C., AND JOVANOV, E. 2006. System architecture of a wireless body area sensor network for ubiquitous health monitoring. *Journal of Mobile Multimedia 1,* 4.

PASCHALIDIS, I. C. AND GUO, D. 2009. Robust and distributed stochastic localization in sensor networks: Theory and experimental results. *ACM Trans. Sensor Networks 5,* 4, 34:1–34:22.

QUWAIDER, M. AND BISWAS, S. 2008. Body posture identification using hidden Markov model with a wearable sensor network. In *Proceedings of the ICST 3rd International Conference on Body Area Networks*. Brussels, Belgium.

RAY, S., LAI, W., AND PASCHALIDIS, I. C. 2006. Statistical location detection with sensor networks. *Joint special issue IEEE/ACM Trans. Networking and IEEE Trans. Information Theory 52,* 6, 2670–2683.

TORFS, T., LEONOV, V., HOOF, C. V., AND GYSELINCKX, B. 2007. Body-heat powered autonomous pulse oximeter. In *Proceedings of the 5th IEEE Conference on Sensors*. 427–430.