Multi-target Classification and Tracking using Wireless Sensor Networks

A Thesis

by

JOONGHYUN LEE

Presented to the Faculty of the Graduate School of
Seoul National University
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

School of Mechanical & Aerospace Engineering

Seoul National University

Supervisor: Professor H. Jin Kim

February 2015

to my

MOTHER, FATHER, and SISTER

with love

Abstract

This thesis presents an acoustic classification and multi-sensor tracking algorithm with WSNs for multiple targets is suggested. The goal for this study is to classify and track the traces of moving multi-targets with their acoustic characteristics and received signal strength indicator. The thesis includes unique method to select features from the raw acoustic signal which contains both time and frequency domain analysis. For localization, Gaussian process based algorithm for estimating unidentified traces from received signal strength indicator is presented. In addition, the unique method for labeling unidentified traces with appropriate target is introduced. By using suggested algorithm, the classifier shows more accurate and faster responding performance than the classifiers which use only frequency domain input. Experimental results with show the satisfactory performance of the proposed algorithm.

keywords: multi-target tracking, acoustic classification, wireless sensor networks, SVM classifier, Gaussian process based localization

student number: 2013-20701

Table of Contents

List of Figures

List of Tables

1

Introduction

Wireless sensor networks (WSNs) consist of small and low-power sensor nodes that collect, process and communicate the specified physical state. The WSNs are appropriate for monitoring environmental phenomena that are widely distributed in space and time. This paper focuses on classification and tracking of multiple targets moving within the region of interest where sensor nodes are deployed. Classification and tracking of multiple targets has an important role in military applications [1], habitat monitoring [2] and traffic monitoring [3] in surveillance system. In particular, this paper considers the acoustic signal that is widely used for classification since it offers a rich set of features.

1.1 Literature Review

Once set of features are extracted, it can be used to classify acoustic signals by conventional classifier such as support vector machine (SVM) [4], k nearest neighbor rule (KNN) [5], multi-layer perceptron (MLP) [6] and artificial neural networks (ANNs) [7].

In this study, the method to extract features from the raw signal is one of the main concerns for classification part. There are several methods to analyze the acoustic data. The most conventional tool of signal analysis is Fourier transform (FT). FT converts time domain signal into constituent sinusoids of different frequencies. However, FT has a disadvantage in that it loses the time domain information of the signal. For example, analyst cannot tell when a particular event took place from a Fourier Spectrum. Especially for a non-stationary signal, time domain analysis is not possible through FT.

To overcome this drawback, short time fourier transform (STFT), using a technique called windowing, was proposed [8]. STFT maps the signal into a two-dimensional space of time and frequency using single fixed window. STFT can be regarded as a compromise between the time information and frequency information. It provides some information about both time and frequency domain. However, the precision of the information is limited by the size of the window. The size of windowing function which means length of data sample determines whether there is good time resolution or good frequency resolution. Good frequency resolution separates the frequency component into small pieces. Good time resolution represents the time at which frequencies change. A wider window gives better frequency resolution but poorer time resolution and vice versa

Wavelet transform (WT) represents the next logical step: a windowing technique with variable size. Thus, it preserves both time and frequency information of the signal. Wavelet techniques are successfully applied to various problems in signal and image processing. Data compression [9], motion estimation [10], segmentation and classification [11, 12] and denoising [13] are only some examples. It is perceived that the wavelet transform is an important tool for analysis and processing of signals and images. In spite of its efficient computational algorithm, the wavelet transform suffers from two disadvantages that undermines its application for certain signal and image processing tasks [14, 15]. They are shift sensitivity and directional selectivity. A transform is shift sensitive, if the shifting in time. For input-signal causes an unpredictable change in transform coefficients. It has been observed that the standard discrete wavelet transformation (DWT) is severely disadvantages by the shift sensitivity that arises from down samplers in the DWT implementation [14, 16]. Shift sensitivity is an undesirable property because it implies that DWT coefficients fail to distinguish between input-signal shifts. Especially the setting that has frequent changes in input-signal, this is crucial defect for feature data that is used to classify acoustic data.

On the other hand, the low cost sensor limits to adopt existing multi-target tracking algorithm. First, it does not satisfy the assumption that the individual observation of the each sensor node has information of the target directly. Based on this assumption, many algorithms such as Gaussian mixture probabilistic hypothesis density filter [17], Markov chain Monte Carlo data association [18] and other research [19, 20] are developed. However these algorithms cannot be applied to the low cost sensor networks. Second, the low-cost sensor node does not use additional hardware support. Conventional tracking algorithm such as tome of arrival (TOA), time difference of arrival (TDOA) and angle of arrival (AOA) need special high cost hardware [21]. For example, the radar sensor used in [22] needs much more electrical power than small-size sensor node such as MicaZ, so the node equipped with the radar sensor has to become huge hardware. Third, the low-cost sensor limits data memory. Some works for target classification described in [23], [24] and [25] use fast Fourier transformation (FFT) and image processing in order to extract feature data from raw data. These need excessive calculation time and large data memory, which are intractable for low-cost sensor node without additional digital signal co-processors.

A few papers adopt the low-cost sensor nodes for the multi-target tracking scenario [26] focuses on classification of different target based on the signal characteristics of employed sensors. [27] defines the specific areas that each specific target is potentially located. The areas are used to classify the target via sensor fusion and approximate the locations of target as the centers of each area. This algorithm needs exact empirical sensor modeling and the tracking allows default error whose maximum is volume of the defined area. From there limitations and the lack of researches, it could be claimed that simultaneous classification and tracking of multi-target is still challengeable for low cost sensor network.

On the other hand, unlabeled data which becomes unidentified traces is not directly obtained from the sensor measurement. It is generated from predictive target distribution which represents the probability of predicted target locations. For instance, if interested locations of the targets are in two dimensional space, the predictive target distribution is 3-D distribution where a value z(x,y) on z-axis shows the probability that the targets are located at the (x,y) point. As the representative method obtaining the predictive target distribution, a Gaussian Process (GP) regression is used in [28]. It gives precise modeling against noisy environment. (picture) predictive target distribution which gives information about the predicted locations of targets in a surveillance area. As the samples for making the predictive target distribution, we employ received signal strength indicator (RSSI) which is nonlinear and noisy measurement. The RSSI is popular choice for the low-cost sensor networks as in [29] [30] and [31] because most commercially available transceivers can process the RSSI. In addition, it does not have a noticeable weak point on power consumption, cost and size of the sensor node.

1.2 Thesis Contribution

The main purpose of this thesis is to propose classifier that includes both frequency and time domain features and introduce labeling algorithm for the unidentified traces.

First, the thesis contains the classifier with acoustic data input and eight different classes as output. The feature set used for classifier is consist of both frequency and time domain which gives both accurate and time sensitive classification performance.

Second, in order to meet the requirement of the integrated classification and target tracking, an algorithm based on estimating distance from the acoustic sensor to each target is proposed. This is referred to as *Trace Labeling* in this work. The distance from the acoustic sensor to each target is calculated from both localization results and acoustic data modeling.

The proposed algorithm is applied to an experiment for classification and tracking of moving one aerial and two ground vehicles. It is hard to compare with other algorithms due to different experimental setting. Instead, I show the experimental results to verify the accuracy and validity of an algorithm. The entire flowchart is depicted below.

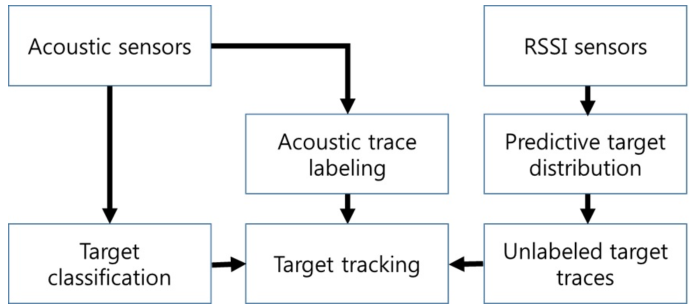

Figure 1.1: Flow chart of whole process

1.3 Thesis Outline

The structure of this paper is organized as follows. A problem formulation and experimental setting are described in chapter 2. In chapter 3, the acoustic classification method is introduced. The classification algorithm contains three parts: 1) feature extraction, 2) feature reduction, 3) regressor construction. SVM regressor is considered for the classifier. The experimental results for performance of classifier is presented at the end of the chapter. Gaussian process based target tracking is shown in chapter 4. The Gaussian process regression is used for estimating unidentified traces. Experimental result is also attached at the end of the chapter. The basic concept and data processing method for trace labeling are introduced in chapter 5. Finally, in chapter 6, conclusions are given.

2

Experimental Setting

In this section, we describe the experimental setting with scheme for overall process. The data process is summarized into three steps : acoustic classification, Gaussian process based target tracking and trace labeling algorithm. Since each part has different features and evaluation The experimental results will be presented and analysed in each chapter. However, whole process share the experimental setting and rather be introduced once in the beginning. In that, this chapter we describe experimental setting.

In this study, we employ a SVM training approach for multi-target classification in WSNs and GMM for target localization. The term *localization* in this paper means estimating the position of target without identification while the term *target tracking* includes identification of target. The main objectives of multi-target classification and target tracking is firstly to classify the types of targets using SVM classifier with acoustic measurement and secondly to estimate the position of targets separately based on RSSI sensor measurements and trace labeling algorithm. In order to validate feasibility and performance of the tracking algorithm, we construct one scenarios as shown in Fig.2.1

In the experimental setup shown in Fig. 2.1, a total of 18 RSSI sensor nodes and two

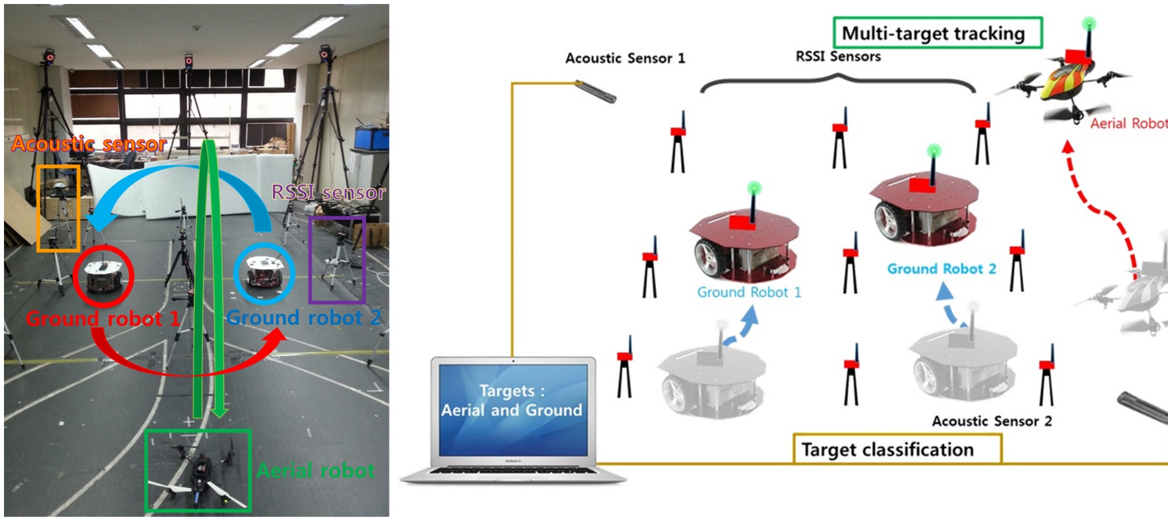

Figure 2.1: Experimental setup for scenario classification and tracking of three heterogeneous targets, two ground robots and an aerial robot.

acoustic sensors are deployed in an indoor environment. All sensors lie on a flat two-dimensional region of interest and two heterogeneous targets (a ground robot and an aerial robot) move in the workspace.

For the experiment, each sensor node employs TinyOS with TelosB platform and communicates using the IEEE 802.15.4 standard. Each node has a micro processor MSP430 and sensors such as received signal strength indicator (RSSI) sensors. Acoustic sensors measure sound pressure level (SPL) and send it to the central processor (PC) using the WiFi. As targets for classifying and tracking, we use two ground robots and one aerial robot which are mobile robots Stella (from NTRexLAB) and a quadrotor ARdrone (from Parrot) respectively. Although the quadrotor is loud enough to make measureable acoustic signal, ground robots are relatively silent which makes measurement difficult. For the sake of experimental convenience, we added acoustic signal of car and jet for two ground robots. The artificially affixed signals are both 30 second long. In this paper, we will call three robots as car, quadrotor and jet respectively. The true trajectories of targets are measured by a VICON vision-based tracking system in order to validate the tracking performance.

3

Acoustic Classification

3.1 Problem Formulation

This subsection describes an algorithmic framework of machine learning for target classification in WSNs. Assume that the measurement differences among the acoustic sensors is negligible in the workspace especially for the classification. Without loss of generality we can use data measurement from the acoustic sensor 1 only for the classification. Since there are two different robots in the workspace, four different type of classes (ground robot, aerial robot, none of them and both of them) are available for acoustic measurement. For the sake of convenience, let four classes : ground robot, aerial robot, none of them and both of them, as A,B,C and D respectively. The acoustic sensor 1 measures local binary decision variables $y^{IJ}(t)$ at time t, where $y^{IJ}(t) = 1$ indicates class I with current situation for $I, J \in [A, B, C, D] \, and \, I \neq J$, while $y^{IJ}(t) = -1$ indicates class J. For the SVM training, the acoustic sensor 1 should generates data in the form of input-output pairs. Considering six types (AB, AC, AD, BC, BD and CD) of classifier, we can obtain the paired data from the acoustic sensor 1 in the form of $M_1(t) = \left(\bar{x}_1(t), y_1^{(IJ)}(t)\right)$ for k= AB, AC, AD, BC, BD and CD with time t. The input vector can be defined as $\bar{x}_1 = (x_{fd1}, x_{td1})$ which x_{fd1} and x_{td1} represent the extracted features from frequency domain and time domain respectively. After training with paired data, we can obtain a function $f_{IJ}(\bar{x}_1)$ that discriminates which class the acoustic data measurement is more likely to be among the AB, AC, AD, BC, BD and CD.

Suppose that there are binary decision variables y^{IJ} for $I, J \in [A, B, C, D]\, and\, I \neq J$. The data fusion center gathers the measured data from each sensor node and solve the following dual form of the SVM problem for each type of classifier.

$$\min_{\lambda_i{}^{(IJ)}} \frac{1}{2} \sum_{i,j} y_i{}^{(IJ)} y_j{}^{(IJ)} \lambda_i{}^{(IJ)} \lambda_j{}^{(IJ)} \Phi\left(\bar{x}_i, \bar{x}_j\right) - \sum_i \lambda_i^{IJ} \tag{3.1}$$

$$s.t. \sum_i \lambda_i^{IJ} y_i^{IJ} = 0, 0 \leq \lambda_i^{IJ} \leq C, i, j = 1, \cdots, L\,, \tag{3.2}$$

where, $\left(\bar{x}_i, y_i^{(IJ)}\right)$ is an input-output pair, λ_i^{IJ} is the corresponding Lagarangian multiplier for $I, J \in [A, B, C, D]\, and\, I \neq J$, C is a regularization constant, L is the length of training data, and $\Phi\left(\cdot, \cdot\right)$ is a kernel function to express the nonlinearity of the data distribution. Solving the optimization problem (1) with quadratic programming, we can obtain the discriminant function as,

$$f_{IJ}(\bar{x}) = \sum_{j=1}^{L} y_j{}^{(IJ)} \lambda_j{}^{(IJ)} \Phi\left(\bar{x}, \bar{x}_j\right) + b \tag{3.3}$$

and for each type of classifier such that $\lambda_i^{IJ} > 0$, we call the corresponding input-output pair $\left(\bar{x}_i, y_i^{(IJ)}\right)$ the support vector [32].

3.2 Methodology

Acoustic classification method consists of three stages as shown in Fig 3.1. This is the cardinal principle and standard classification method which has already been proved as the best classification method [33]. Selecting appropriate feature set from raw data is key issue

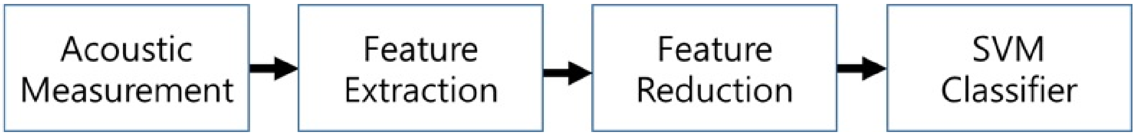

Figure 3.1: Methodology of our proposed algorithm.

for accurate classification. Since the classification task is a multi-stage process, the role of dimensionality reduction is to reserve information that is important for class distinguishing and renounce that which is irrelevant. A classifier with fewer inputs has fewer parameters to be determined, leading to a classifier with better generalization properties.

There are two ways to analyse acoustic signal: frequency and time domain analysis. Since both domain analysis are complementary for each other, we select feature from both domain for SVM inputs. The time domain analysis gives real-time information but contains low dimension of feature set while the frequency domain analysis gives high dimension feature space but loses real-time performance. Unlike to other feature extraction algorithm, we construct feature set from frequency and time domain linearly. The latest technique such as WT uses coupled feature set of time-frequency analysis which causes high computational load and difficulties in parameter tuning.

3.3 Feature Extraction and Reduction

Since it is difficult to use raw acoustic measure directly into SVM classifier, feature extraction and reduction process are essential. For the frequency domain analysis, we select STFT with window size as two seconds. As this window size gets wider, the real-time performance gets poorer. As shown in Fig. 3.2, high dimension feature space is extracted from raw acoustic data. This STFT graph shows unique distribution with different targets. The position and amplitude of peak point cluster varies over different target signals.

For the time domain analysis, we select Weibull likelihood function. Weibull function is a continuous probability distribution which is widely used for indicating failure rate. The probability density function of a Weibull random variable is: $f(x;\lambda,k) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(x/\lambda)^k}$,

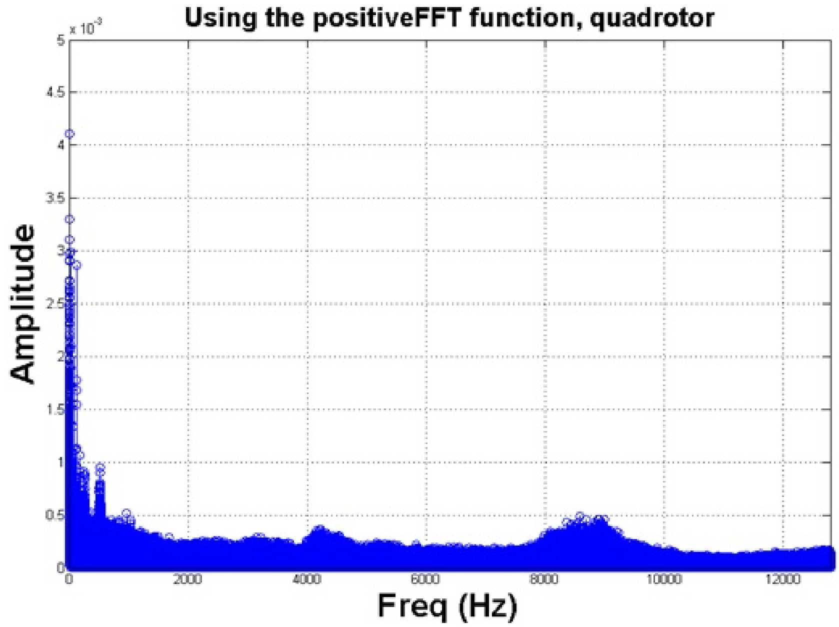

Figure 3.2: Frequency domain analysis : short time Fourier transform of quadrotor sound with window size two seconds.

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. The Weibull likelihood function is the function that calculated the likelihood between input signal and it's Weibull fit function. The Weibull fit function calculate the parameter λ and k which makes the closest Weibull distribution with respect to original signal in root mean square error manner. As a result, Weibull likelihood function returns single scalar value when the input signal is given. Fig. 3.3 represents Weibull likelihood of 4 different cases versus time. As shown in figure, each Weibull likelihood result is distinguishable with different cases. Both time and frequency domain feature space are extracted from the raw acoustic measurement and these features from two domain of feature space are merged into an input vector for SVM input.

Fig. 3.4 and Fig. 3.5 is the verification of feature set for 4 cases of our experimental scenario. As shown in figure, selected feature space are suitable for classification.

When the acoustic sensor measures data, the two seconds of acoustic data is transformed

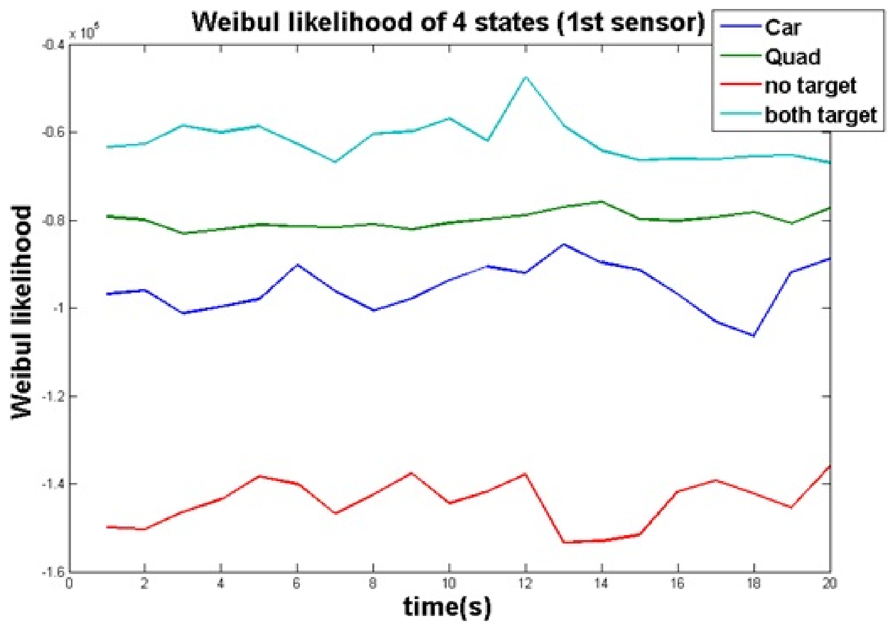

Figure 3.3: Time domain analysis : Weibull likelihood function of 4 cases.

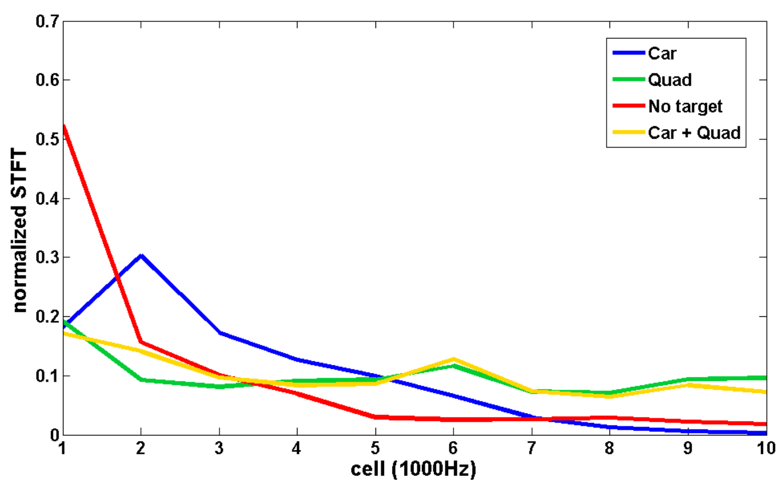

Figure 3.4: Segmented STFT distribution for each 4 case. There are clear differences among the cases.

into STFT distribution. This distribution contains 51.2k feature points with 25.6kHz range. So the STFT coefficients $c_i(t)$ are calculated at every seconds for $i = 1, \cdots, 25600$.

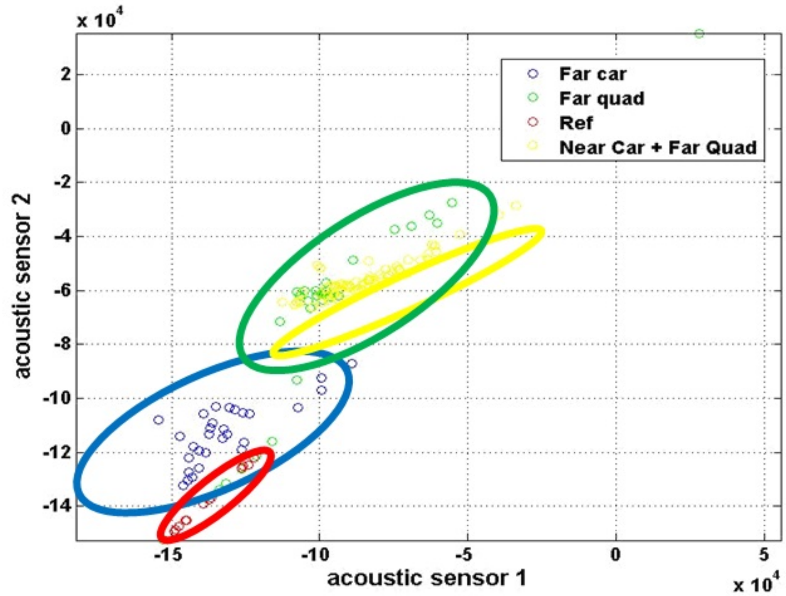

Figure 3.5: Weibull likelihood measured from two acoustic sensor for each 4 case. There are differences among the cases.

For feature redcution, STFT values are segmented with 1kHz from 1Hz to 10kHz and averaged with section. Therefore ten reduced feature points can be represented as following: $C_i(t) = \sum_{j=1+1000i}^{1000+1000i} c_j(t)/1000$ for $i = 0, \cdots, 9$. For normalizing scale of input vector, $C_i(t)$ are normalized. Thus segmented and normalized STFT value $Cnorm_i(t)$ is represented as following: $Cnorm_i(t) = \frac{C_i(t)}{\sum C_i(t)}$. Fig. 3.6 shows the raw feature set and the reduced feature set.

Meanwhile, Weibull likelihood function is calculated with the one second of acoustic measurement input. Therefore, $Cnorm_i(t)$ for $i = 1, \cdots, 10$ and $wbl\,(t)$ are the input vector $\bar{x}(t)$ for every second. For SVM training, labeled data $y^{(IJ)}$ are given from the user. As a result, the input-output pair $\left(\bar{x}(t), y^{(IJ)}(t)\right)$ for $t = 1, \cdots, L$ is constructed. This method of selecting feature set is different from existing method since there is a direct time domain feature space. Fig. 3.7 shows the difference in selecting feature set.

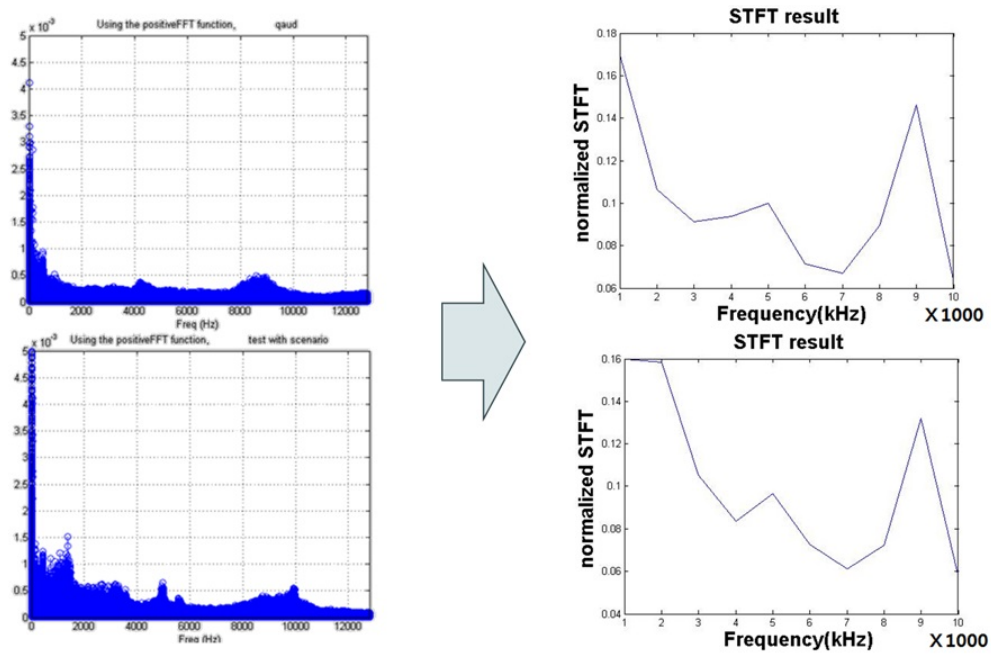

Figure 3.6: The feature space of raw data has so many feature points (left), so reduction of feature set is applied to the feature space (right).

3.4 Classification Algorithm

For accomplishing classification, SVM classifier requires two stages: training and test. In this experimental scenario, there are three different targets which makes $2^3 = 8$ different class of acoustic signal. Thus the acoustic signal of each class is measured and the discriminant function $f_{IJ}(\bar{x}) = \sum_{j=1}^{L} y_j{}^{(IJ)} \lambda_j{}^{(IJ)} \Phi\left(\bar{x}, \bar{x}_j\right) + b$ from the Eq. 3.3 is constructed for each (I, J) pair for training stage. After the training stage, the classification tree can be constructed for investigating the class of unknown acoustic measurement. Fig. 3.8 shows classification tree for 6 types of classifiers with 4 different classes. In this experimental setting, $\binom{8}{2} = 28$ types of classifiers with 8 different classes are constructed. After passing through the classification tree, the class of unknown signal can be categorized.

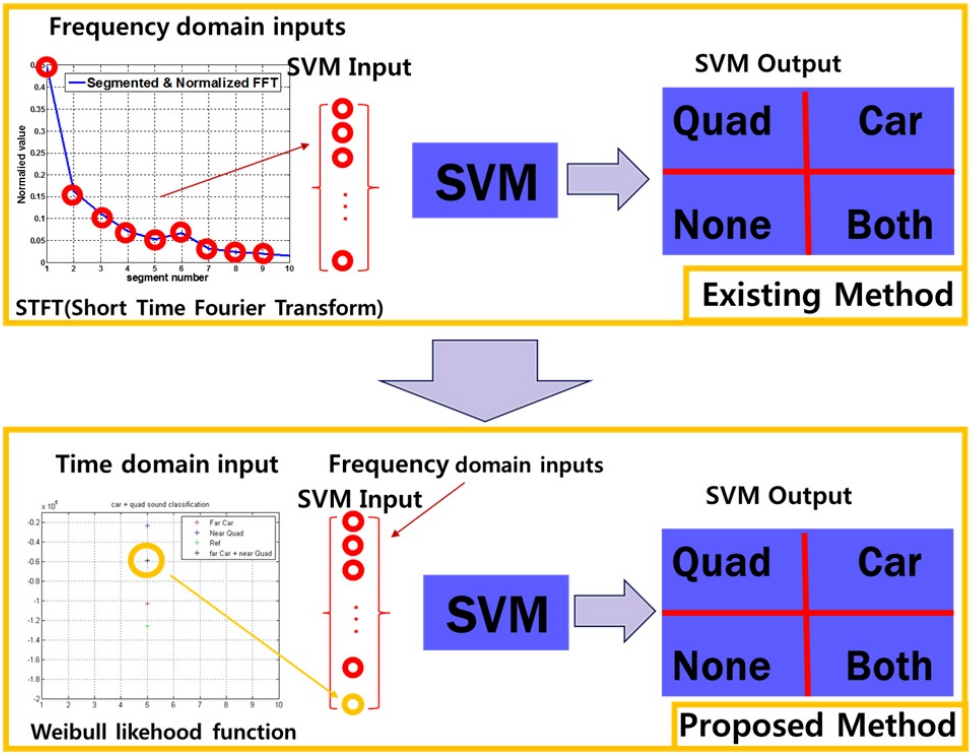

Figure 3.7: Difference between existing method and proposed method in selecting input vector

3.5 Experimental Results

In this section, we describe the results of experimental scenario of WSNs especially for acoustic classification part only. In this experimental scenario, three heterogeneous targets (a car sound, a quadrotor sound and a jet sound) are deployed and turn on and off with 30 seconds of certain pattern. Fig 3.9 shows comparison of proposed classifier and existing classifier with the regressor output as following: 1 for a car, 2 for a quadrotor and 3 for a jet. The blue line is true scenario, the red line is proposed algorithm and the green line is existing algorithm. Since the time domain data is selected as feature set, it shows quicker respond through time. However, there is still some delay caused by STFT window size.

Since quicker respond leads classifier to more accurately, the classification performance

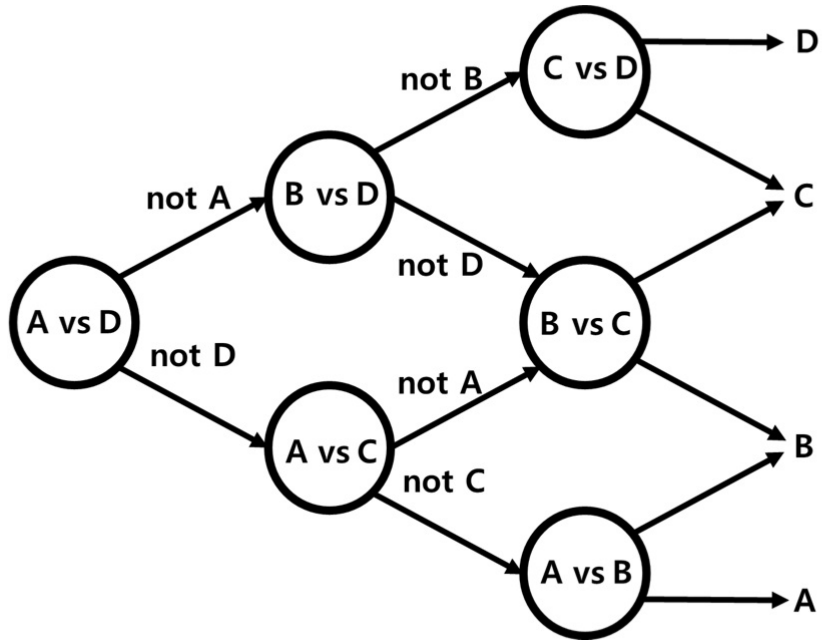

Figure 3.8: Classification tree for 6 types of classifiers with 4 different classes.

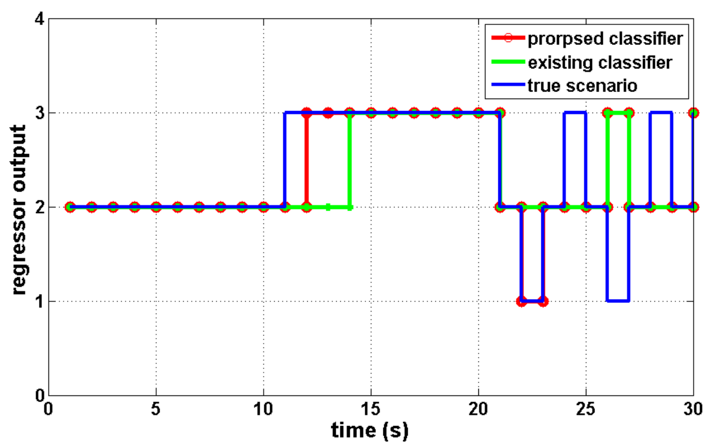

Figure 3.9: Comparison of proposed classifier and existing classifier with the regressor output as following: 1 for a car, 2 for a quadrotor and 3 for a jet.

is also slightly improved with proposed algorithm. Table.3.1 shows the classification succeed rate of existing algorithm and proposed algorithm.

Class	Existing algorithm (STFT only)		Proposed algorithm (+Weibull likelihood)	
	Training data	Test data	Training data	Test data
car	100	100	100	100
quad	100	86.15	100	96.87
no target	96.87	96.87	96.87	100
car+qaud	95.24	86.75	100	92.75
jet	100	100	100	100
jet+car	100	94.08	100	97.82
jet+quad	96.87	95.72	98.63	97.82
all of them	92.57	90.53	95.33	93.14

Table 3.1: Classification success rate comparison of existing algorithm and proposed algorithm for 8 different classes

4

Gaussian Process based Localization

In this section, we describe the localization algorithm with Gaussian process. As the samples for constructing the predictive target distribution, we employ RSSI sensor which is nonlinear and noisy measurement. To overcome the nonlinearity and noise, probability based approach is applied for estimation. The goal for this part is to estimate the traces of multi-targets moving in the region of interest.

4.1 Problem Formulation

This subsection briefly describes the method for constructing predictive target distribution. In a RSSI-based aproach, the signal strength emitted by targets is measured by several distributed sensors near the target. With the assumption that the location of sensor nodes are known, these measurements are processed into 3-D distribution which represents the probability of the predicted target locations. Let this distribution as *predictive target distribution*.

In order to construct the predictive target distribution, we use radial basis function (RBF) based approximation. Let N units of RSSI sensors measures RSS signal from target and the true target position as $X = [x, y]^T \in R^2$, position of i_{th} sensor node as $S_i = [s_{i1}, s_{i2}]^T \in R^2$, measurement of i_{th} sensor as $y_i \in R$ and y_{Ri} is the true RSS signal. Then RSS measurement can be represented as following:

$$y_i = y_{Ri} + v_i = p_0 - 10\alpha \log(\| X - S_i \| + v_i) \tag{4.1}$$

p_0 means free-space Friis model [34] constant which is RSS measurement at the unit reference distance. α is path loss exponent constant and v_i is the Gaussian measurement noise.

Let the position and measurement vector of sensor nodes as $\bar{S} = [S_1, \cdots, S_N]$, $Y = Y_R + V = [y_{R1}, \cdots, y_{RN}] + [v_{R1}, \cdots, v_{RN}]$. The RSSI sensor measurement with noise can be represent as following equation.

$$Y \sim N(Y_R, K(\bar{S}, \bar{S}) + \sigma^2 I_{n\times n}), \tag{4.2}$$

where Y is the measured RSSI sensor data, $K(\bar{S}, \bar{S})$ is the kernel function matrix with $(i, j)_{th}$ element is $k(\bar{s}_i, \bar{s}_j)$. In this paper, kernel function is defined as $k(\bar{s}_i, \bar{s}_j) = \theta_1 exp(\frac{-\|s_i - s_j\|^2}{2\theta_2})$. The parameter θ_1, θ_2 are estimated with maximum likelihood using conjugate gradient method. Note that this kernel function is different kernel function from SVM dual problem. Through the Gaussian Process Regression, with the X which is the position of target, the expected RSS measurement and the standard variation can be calculated. Figure 4.1 is the example for 3-D distribution of predicted location of the targets.

4.2 Gaussian Process Regression

In order to make the predictive target distribution described in section 4.1, we utilize RBF based approximation. As the representitive method obtaining the predicted target distribution, a Gaussian process regression works well in modeling environmental information described in [35]. Derived from equation. 4.2, expected RSS intensity $\hat{y}(X)$ and it's variance $\sigma_{gp}^2(X)$ of unknown target position X is represented as following :

$$\hat{y}(X) = K\left(X, \bar{S}\right)^T (K(\bar{S}, \bar{S}) + \sigma^2 I_{n\times n})^{-1} Y \tag{4.3}$$

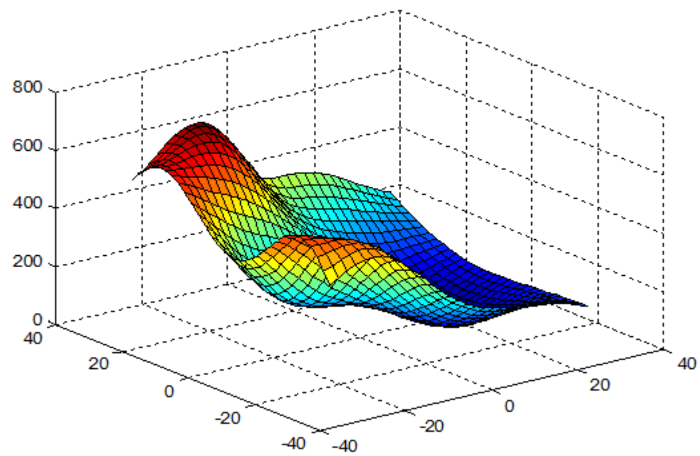

Figure 4.1: The 3-D distribution using the radial basis function (RBF) based approximation. In this distribution, higher value on z-axis represents higher probability of the predicted locations of the targets.

$$\sigma_{gp}^2(X) = K\left(X, \bar{S}\right)^T \left(K(\bar{S}, \bar{S}) + \sigma^2 I_{n\times n}\right)^{-1} K\left(X, \bar{S}\right), \tag{4.4}$$

where position and measurement vector of sensor nodes as $\bar{S} = [S_1, \cdots, S_N]$, $Y = Y_R + V = [y_{R1}, \cdots, y_{RN}] + [v_{R1}, \cdots, v_{RN}]$, $K(\bar{S}, \bar{S})$ is the kernel function matrix with $(i,j)_{th}$ element is $k(\bar{s}_i, \bar{s}_j)$. In this paper, kernel function is defined as $k(\bar{s}_i, \bar{s}_j) = \theta_1 exp(\frac{-\|s_i - s_j\|^2}{2\theta_2})$. Using RSS measurement, the 3-D distribution of targets are predicted. Fig. 4.1 shows two peaks of target distribution. These peaks can be regarded as the estimated position of each target.

4.3 Experimental Results

In this subsection, localization result of experimental scenario described in Sec. 4.2 is presented. In this experimental scenario, we deal with multi-target tracking problem in a 2-dimensional workspace $W = [(x_1, x_2)| -250cm \le x_1 \le 250cm, -250cm \le x_2 \le 250cm]$.

Fig. 4.2 shows the localization results. Dashed lines are estimated position of each target and solid lines are true position measured from VICON. Purple dots are the acoustic sensor

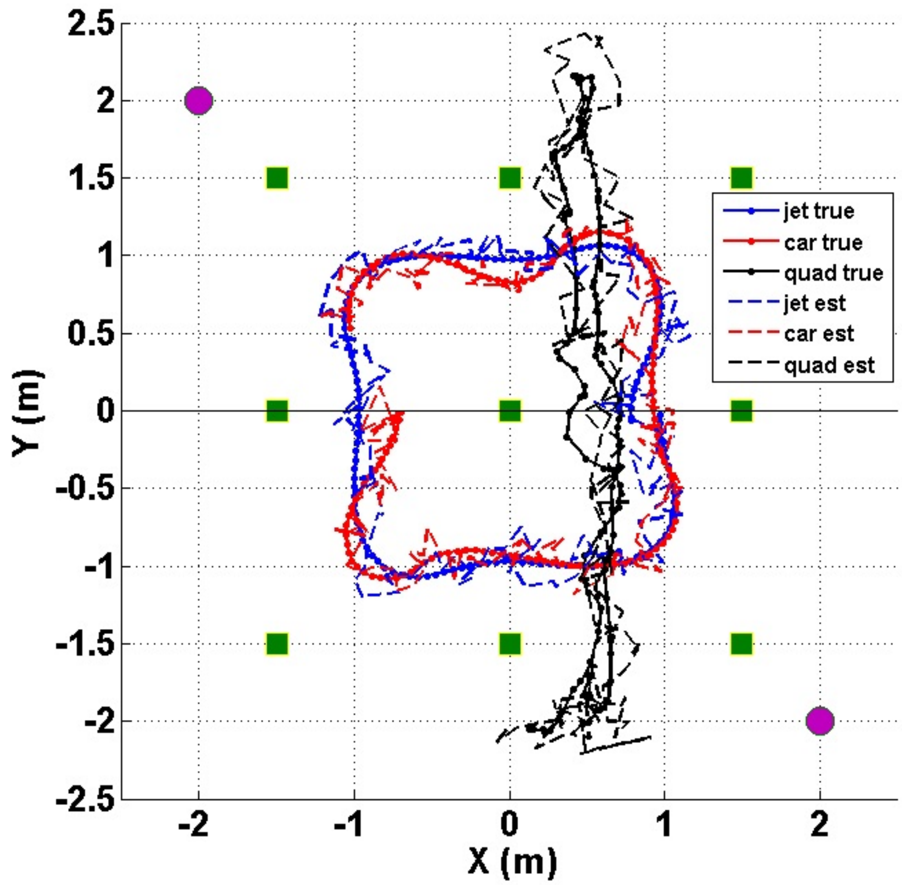

Figure 4.2: Localization results for three moving targets: car, quadrotor and jet. Purple dots are the acoustic sensor and green dots are the RSSI sensors.

and green dots are the RSSI sensors. Although there is no labels for traces in this steps, we colored the traces with appropriate target for the convenience in visualization manner.

The calculated average root mean-squared error (RMSE) of the ground vehicle is 20.38cm and 34.72 for aerial vehicle. We infer that the moving speed and dynamics of the two targets causes the difference in RMSE. Since the aerial vehicle moves faster and noisier than the ground robot, the RMSE of aerial vehicle is slightly larger than ground vehicle.

5

Data Integration with Trace Labeling

In this section, we suggest the trace labeling algorithm that designates the unidentified traces with the relevant targets. The goal of this part is that identify the traces by comparing distance from acoustic sensor to each target. The time series of distance data from localization part and acoustic modelling part can be compared with convolution method. In order to meet the requirement of the integrated target tracking, the labeling process should not be post-processing. However, smaller window size of distance data makes lower labeling performance accuracy. Therefore, there is a compromise in labeling accuracy and time delay.

5.1 Problem Formulation

For the last part of multi-target tracking, we introduce trace labeling as matching algorithm. The trace labeling is the process that labels the unidentified traces from localization part to the appropriate targets. Since each target makes certain trace and it's own acoustic signature, it is possible to designate the traces with the relevant targets. After the localization part, the estimated traces are given but they are unidentified. With the assumption that the position of acoustic sensors is known, we can calculate the distance between target and acoustic sensors for each target. However, there is no information about the identification of targets in this step. Thus, we suggest temporal labeling such as target a,b and c instead of target 1,2 and 3 which represent ground robot 1, aerial robot and ground robot 2 respectively. Then the distance between a target and j_{th} acoustic sensor at time t can be defined as $D_{aj}(t)$ and same form for other targets. On the other hand, the we can also calculate the fractional ratio of each target from acoustic measurement. With the assumptions that the acoustic signal has linearly additive property, region of interest is small enough to make TDOA as zero and the effect of reverberation is negligible, the acoustic measurement can be present as analytic model.

$$X_j(t) = \sum_{i=1}^{n} \frac{a_{ij}}{r_{ij}^2(t)} S_i(t) \tag{5.1}$$

Eq. 5.1 is the acoustic measurement model with fractional ratio where $X_j(t)$ means acoustic measurement of j_{th} acoustic sensor at time t. a_{ij} means acoustic constant for i_{th} target at j_{th} acoustic sensor. $r_{ij}(t)$ is the distance between i_{th} target and j_{th} acoustic sensor at time t. $S_i(t)$ is the acoustic data sample of i_{th} target which is consist of acoustic data of i_{th} target with the 1 second length. The acoustic data sample is prepared before the test scenario by acoustic recording. In this scenario, there are two acoustic sensors and three targets. Therefore i varies from 1 to 3 and j varies from 1 to 2.

After solving the Eq. 5.1, $r_{ij}(t)$, the distance between i_{th} target and j_{th} acoustic sensor at time t can be calculated. In this time, we know the target identification with the estimated distance from the sensor. Let the distance between i_{th} target and j_{th} acoustic sensor at time t as $D_{ij}(t)$. Then it is possible to compare $D_{aj}(t)$, $D_{bj}(t)$ and $D_{cj}(t)$ with $D_{ij}(t)$ for i = 1,2,3. After the comparison, we can finally match the unidentified traces with relevant target identification.

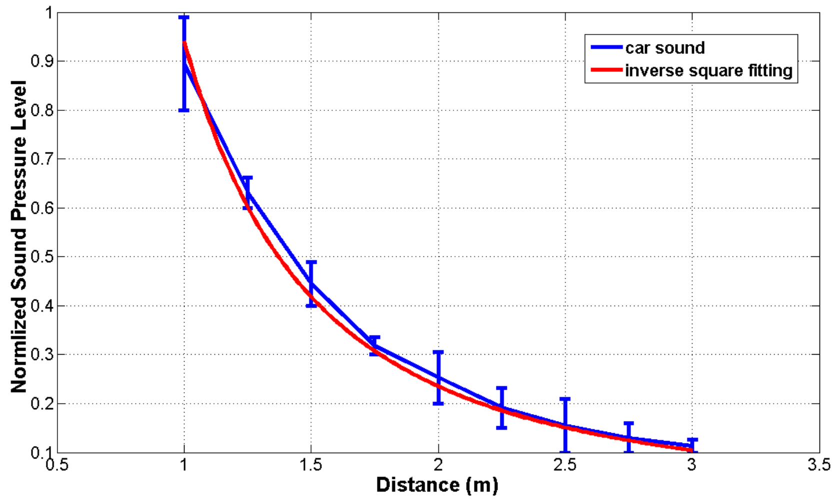

Figure 5.1: Mean and variance of car sound pressure level versus distance from the sound source.

5.2 Acoustic Signal Modeling

With the assumptions that the acoustic signal has superpositive property, workspace is small enough to make TDOA as almost zero and the effect of reverberation is negliable, the acoustic signal measurement can be modelled as $X_j(t) = \sum_{i=1}^{n} \frac{a_{ij}}{r_{ij}^2(t)} S_i(t)$ (Eq. 5.1). The acoustic data sample of each target $S_i(t)$ from Eq. 5.1 is secured when the training for acoustic classification is done. In the acoustic model, each signal component follows inverse square law which means the intensity of signal is inversely proportional to the square of the distance from the source. For the short range of distance, the inverse square property is roughly effective. After investigating simple experiment which moves single target from 1m to 3m, we obtain Fig. 5.1, 5.2and 5.3. The figures show mean and variance of sound pressure level (SPL) for each target versus distance from the source with inverse square fitting. SPL constant a_{ij} from Eq. 5.1 are also calculated with the experiment. Equivalent amount of 95.2 dB, 97.4 dB and 88.4 dB of SPL are estimated as a_{11},a_{21} and a_{31} respectively.

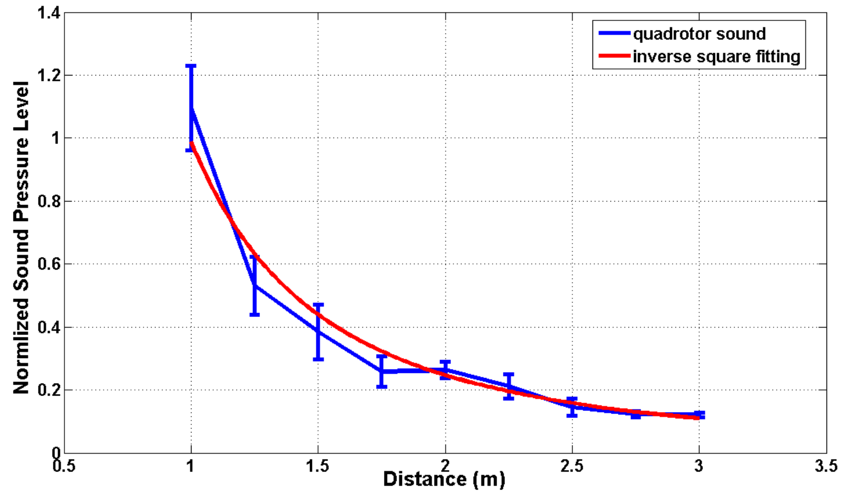

Figure 5.2: Mean and variance of quadrotor sound pressure level versus distance from the sound source.

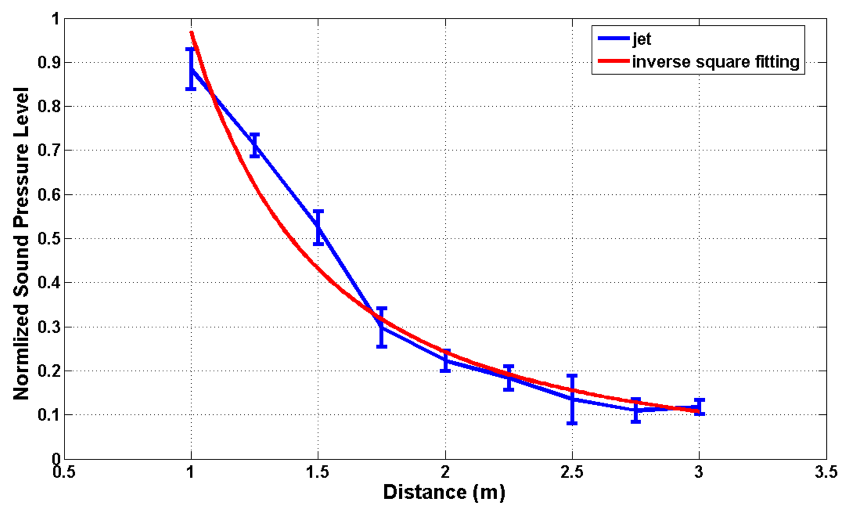

Figure 5.3: Mean and variance of jet sound pressure level versus distance from the sound source.

5.3 Acoustic Data Sampling

Since the raw acoustic signal changes over time very frequently, some transformation is required to process acoustic signal. One way to transform acoustic data is FT which makes the raw signal relatively static data. Apply Fourier transfromation to Eq. 5.1 for each side of equation, we obtained the coefficient equation as following:

$$C_j(t) = \sum_{i=1}^{n} \frac{a_{ij}}{r_{ij}^2(t)} D_i(t), \tag{5.2}$$

where, $C_j(t) = \begin{bmatrix} c_{j1}(t) \\ \cdots \\ c_{jn}(t) \end{bmatrix} = FT(X_j(t))$, $D_i(t) = \begin{bmatrix} d_{i1}(t) \\ \cdots \\ d_{in}(t) \end{bmatrix} = FT(S_i(t))$, and n is half of sampling frequency which is equivalent to 25600/s.

Since this FT coefficient matrix $C_j(t) =$ is calculated every one second, feature reduction is required in this part too. To reduce computational cost, we downsampled FT coefficient from 25600 to 256. Same method as feature reduction in classification part is applied: segmentation and normalization. Fig. 5.4 shows the original and reduced FT distribution of car sound sample. In the same way, acoustic data sample of quadrotor and jet are prepared. Fig. 5.5 and 5.6 shows reduced FT distribution of quadrotor and jet respectively.

5.4 Trace Labeling

For each acoustic sensor j, every time t Eq. 5.2 becomes systems of 256 linear equations with 3 variables $\frac{1}{r_{ij}^2(t)}$ for $i = 1, 2$ and 3. Solving the systems of linear equation by least square method, we can estimated the distance between the acoustic sensor and the each target for every time t. Fig. 5.8 shows estimated fraction of each acoustic signal within experimental scenario described in Sec. 4.2. and Fig. 5.9 shows estimated distance from acoustic sensor to each target. Dashed lines are estimated position of each target from acoustic analysis and solid lines are estimated position calculated from localization part. Although there is no labels for traces from localization in this steps, we colored the traces with appropriate target for the convenience in visualization manner.

On the other hand, with the assumption that position of every sensor node is known, the distance from acoustic sensor 1 to each target can be calculated by using localization

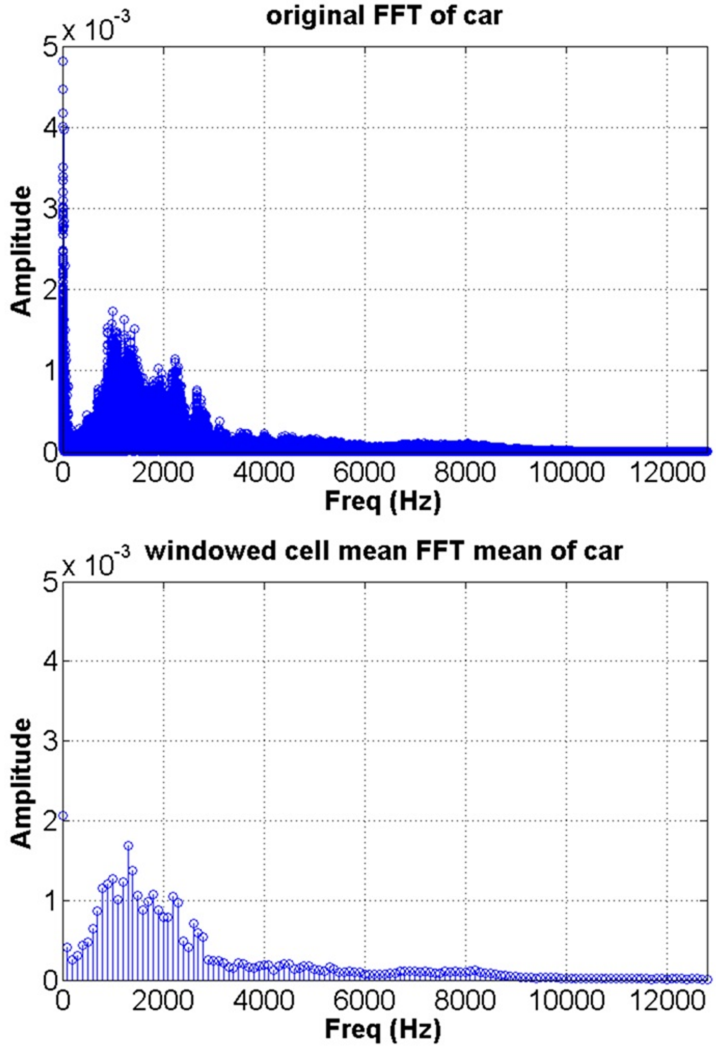

Figure 5.4: Original and downsampled FT distribution. Original FT has 25600 feature points while reduced FT has onl 256 feature points

results. Fig. 5.7 shows the estimated distance from acoustic sensor 1 to each target.

There are two sets of three signals for each part. For localization part, let the distance between i_{th} target and 1_{st} acoustic sensor at time t as $D_{i1}(t)$. Then it is possible to compare $D_{car1}(t)$, $D_{quad1}(t)$ and $D_{jet1}(t)$ with $D_{i1}(t)$ for i = 1,2,3. Since only one to one corresponding relation is possible for this comparison, six combinations are possible:

The average cross correlation computed from each pair, $[D_{car1}(t), D_{11}(t)]$, $[D_{quad1}(t), D_{21}(t)]$, $[D_{jet1}(t), D_{31}$

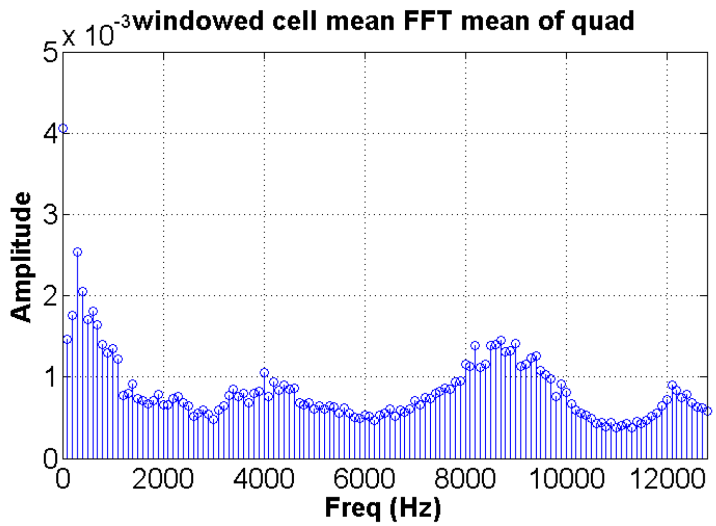

Figure 5.5: Downsampled FT distribution of quadrotor sound.

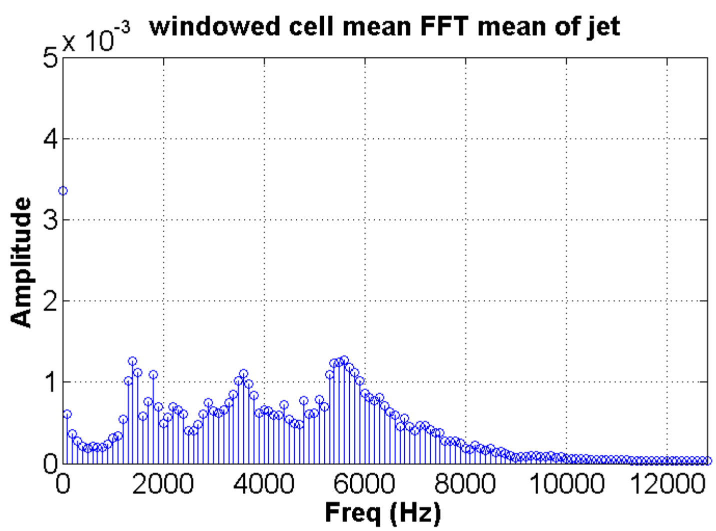

Figure 5.6: Downsampled FT distribution of jet sound.

is highest for the first combination which is correct answer. Therefore the trace labeling algortihm is validated with experimental scenario. After calculating maximum correlation value between distance estimation, the results indicates the appropriate target accurately. However this is the post-processing with whole size window which is not suitable for real-time application. By calculating the labeling accuracy versus correlation window size, we get obtain this Fig. 5.10. From this graph, it is inferred that 10 seconds of window size is

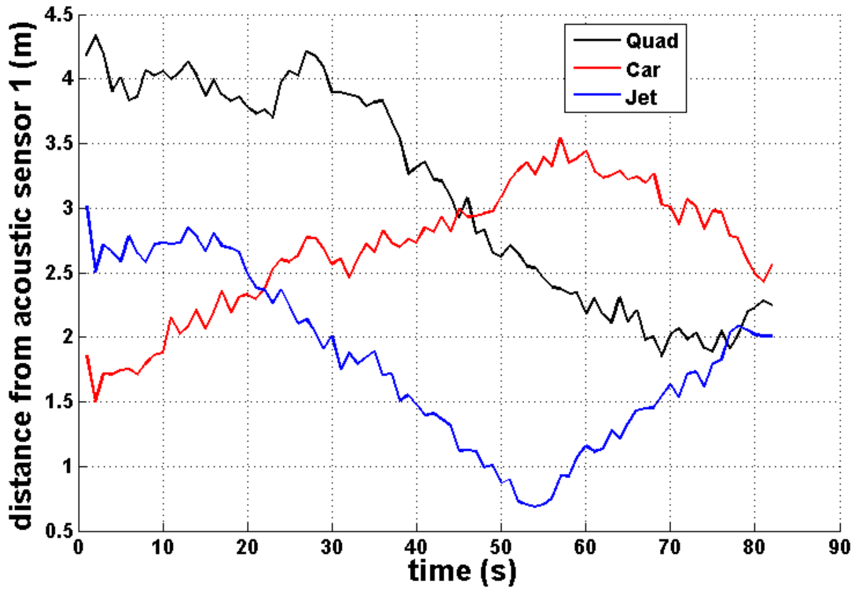

Figure 5.7: Calculated distance from acoustic sensor 1 to each target using localization results.

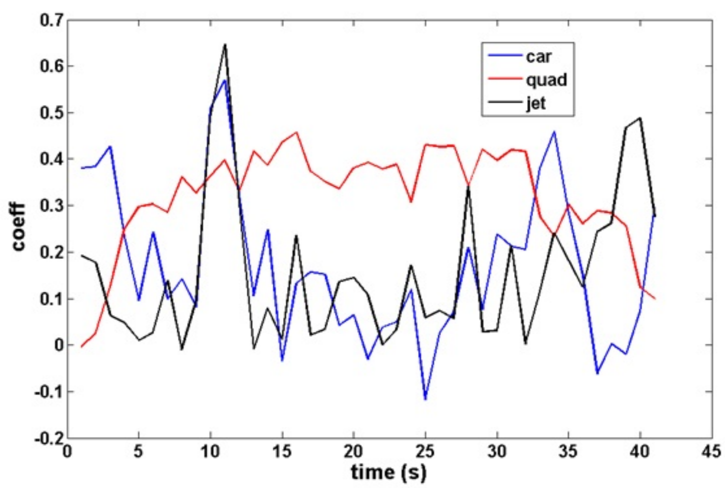

Figure 5.8: Estimated fractional ratio of each acoustic signal within experimental scenario.

enough for successful labeling.

combination	mean cross-correlation
$[D_{car1}(t), D_{11}(t)], [D_{quad1}(t), D_{21}(t)], [D_{jet1}(t), D_{31}(t)]$	0.978
$[D_{car1}(t), D_{11}(t)], [D_{quad1}(t), D_{31}(t)], [D_{jet1}(t), D_{21}(t)]$	0.921
$[D_{car1}(t), D_{21}(t)], [D_{quad1}(t), D_{11}(t)], [D_{jet1}(t), D_{31}(t)]$	0.899
$[D_{car1}(t), D_{21}(t)], [D_{quad1}(t), D_{31}(t)], [D_{jet1}(t), D_{11}(t)]$	0.895
$[D_{car1}(t), D_{31}(t)], [D_{quad1}(t), D_{21}(t)], [D_{jet1}(t), D_{11}(t)]$	0.903
$[D_{car1}(t), D_{31}(t)], [D_{quad1}(t), D_{11}(t)], [D_{jet1}(t), D_{21}(t)]$	0.956

Table 5.1: The average cross correlation computed from each pair is highest for the first combination.

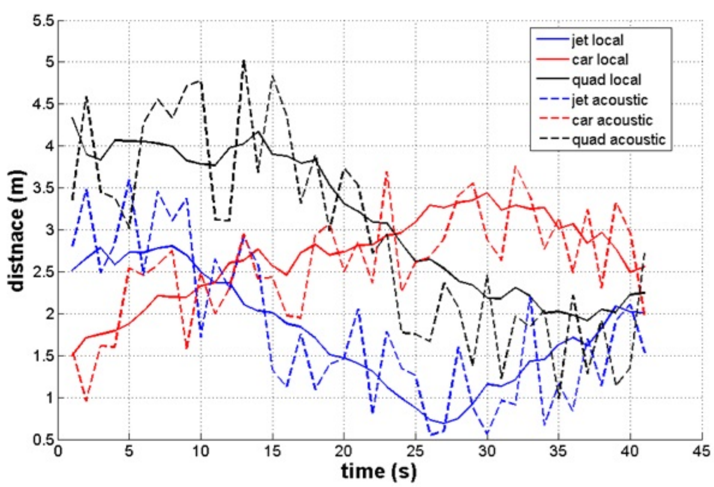

Figure 5.9: Estimated distance from acoustic sensor 1 to each target. .

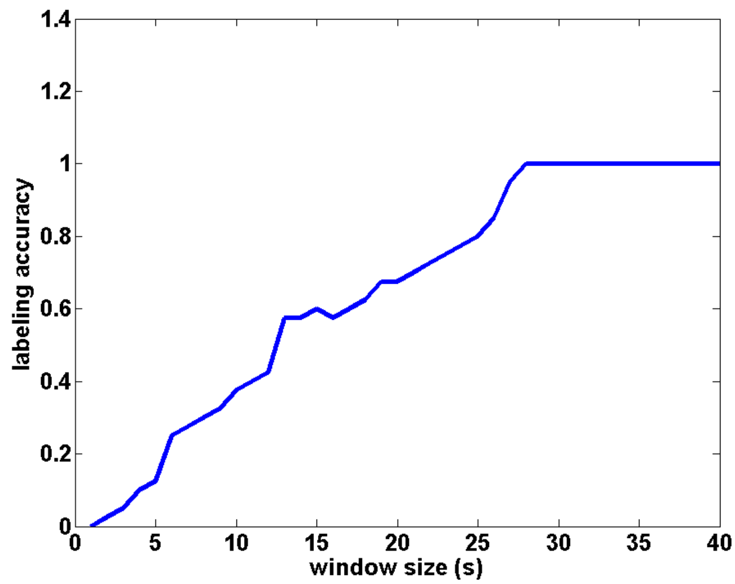

Figure 5.10: Labeling accuracy versus window size.

6

Conclusion

In this thesis, acoustic classification and multi-sensor tracking algorithm with WSNs for multiple targets is suggested. The classifier with both time and frequency domain input for SVM classier makes not only more accurate but also faster response classification performance than classifier which uses only frequency domain input. Also, the labeling algorithm which matches unidentified traces into appropriate target is described. Trace labeling algorithm works well with a certain range of correlation window size. The experimental results show satisfactory classification and tracking performance with three heterogeneous moving targets.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications.* ACM, 2002, pp. 88–97.

[3] X. Li, W. Shu, M. Li, H.-Y. Huang, P.-E. Luo, and M.-Y. Wu, "Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 4, pp. 1647–1653, 2009.

[4] G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machines," *Neural Networks, IEEE Transactions on*, vol. 14, no. 1, pp. 209–215, 2003.

[5] J.-C. Wang, J.-F. Wang, K. W. He, and C.-S. Hsu, "Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio low-level descriptor," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on.* IEEE, 2006, pp. 1731–1735.

[6] K. Englehart, B. Hudgins, P. A. Parker, and M. Stevenson, "Classification of the myoelectric signal using time-frequency based representations," *Medical engineering & physics*, vol. 21, no. 6, pp. 431–438, 1999.

[7] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," *Computers in Biology and Medicine*, vol. 34, no. 6, pp. 523–537, 2004.

[8] J. Allen, "Short-term spectral analysis, and modification by discrete fourier transform," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 25, pp. 235–238, 1977.

[9] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms i," *Communications on pure and applied mathematics*, vol. 44, no. 2, pp. 141–183, 1991.

[10] C. P. Bernard, "Discrete wavelet analysis for fast optic flow computation," *Applied and Computational Harmonic Analysis*, vol. 11, no. 1, pp. 32–63, 2001.

[11] M. Unser, "Texture classification and segmentation using wavelet frames," *Image Processing, IEEE Transactions on*, vol. 4, no. 11, pp. 1549–1560, 1995.

[12] J. Neumann and G. Steidl, "Dual-tree complex wavelet transform in the frequency domain and an application to signal classification," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 3, no. 01, pp. 43–65, 2005.

[13] D. L. Donoho, "De-noising by soft-thresholding," *Information Theory, IEEE Transactions on*, vol. 41, no. 3, pp. 613–627, 1995.

[14] P. D. Shukla, "Complex wavelet transforms and their applications," *Master of Philosophy (M. Phil.) report, University of Strathclyde Glasgow, Scotland*, 2003.

[15] F. C. Fernandes, "Directional, shift-insensitive, complex wavelet transforms with controllable redundancy," Ph.D. dissertation, Rice University, 2002.

[16] G. Strang and T. Nguyen, *Wavelets and filter banks.* SIAM, 1996.

[17] B.-N. Vo and W.-K. Ma, "The gaussian mixture probability hypothesis density filter," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4091–4104, 2006.

[18] S. Oh, S. Russell, and S. Sastry, "Markov chain monte carlo data association for general multiple-target tracking problems," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 1. IEEE, 2004, pp. 735–742.

[19] J. Shin, N. Lee, S. Thrun, and L. Guibas, "Lazy inference on object identities in wireless sensor networks," in *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 2005, p. 23.

[20] J. Liu, M. Chu, and J. E. Reich, "Multitarget tracking in distributed sensor networks," *Signal Processing Magazine, IEEE*, vol. 24, no. 3, pp. 36–46, 2007.

[21] J. Bachrach and C. Taylor, "Localization in sensor networks," *Handbook of sensor networks: Algorithms and Architectures*, vol. 1, 2005.

[22] T. Vercauteren, D. Guo, and X. Wang, "Joint multiple target tracking and classification in collaborative sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 714–723, 2005.

[23] I. Bilik, J. Tabrikian, and A. Cohen, "Gmm-based target classification for ground surveillance doppler radar," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 42, no. 1, pp. 267–278, 2006.

[24] M. F. Duarte and Y. Hen Hu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826–838, 2004.

[25] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*. IEEE, 1998, pp. 8–14.

[26] L. Gu, D. Jia, P. Vicaire, T. Yan, L. Luo, A. Tirumala, Q. Cao, T. He, J. A. Stankovic, T. Abdelzaher *et al.*, "Lightweight detection and classification for wireless sensor networks in realistic environments," in *Proceedings of the 3rd international conference on Embedded networked sensor systems*. ACM, 2005, pp. 205–217.

[27] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda *et al.*, “A line in the sand: a wireless sensor network for target detection, classification, and tracking,” *Computer Networks*, vol. 46, no. 5, pp. 605–634, 2004.

[28] O. Chapelle, B. Schölkopf, A. Zien *et al.*, *Semi-supervised learning.* MIT press Cambridge, 2006, vol. 2.

[29] M. Sugano, T. Kawazoe, Y. Ohta, and M. Murata, “Indoor localization system using rssi measurement of wireless sensor network based on zigbee standard,” *Target*, vol. 538, p. 050, 2006.

[30] F. Viani, L. Lizzi, P. Rocca, M. Benedetti, M. Donelli, and A. Massa, “Object tracking through rssi measurements in wireless sensor networks,” *Electronics Letters*, vol. 44, no. 10, pp. 653–654, 2008.

[31] H. Guo, K.-S. Low, and H.-A. Nguyen, “Optimizing the localization of a wireless sensor network in real time based on a low-cost microcontroller,” *Industrial Electronics, IEEE Transactions on*, vol. 58, no. 3, pp. 741–749, 2011.

[32] V. N. Vapnik, “An overview of statistical learning theory,” *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 988–999, 1999.

[33] A. Ghosh, B. Uma Shankar, and S. K. Meher, “A novel approach to neuro-fuzzy classification,” *Neural Networks*, vol. 22, no. 1, pp. 100–109, 2009.

[34] M. Sato, T. Ohkubo, and H. Niitsuma, “Cross-polarization borehole radar measurements with a slot antenna,” *Journal of Applied Geophysics*, vol. 33, no. 1, pp. 53–61, 1995.

[35] Y. Xu, J. Choi, and S. Oh, “Mobile sensor network navigation using gaussian processes with truncated observations,” *Robotics, IEEE Transactions on*, vol. 27, no. 6, pp. 1118–1131, 2011.

국 문 초 록

본 논문에서는 무선 센서 네트워크를 이용한 다중 물체의 음향을 통한 분류와 다중 센서 위치 추정 알고리즘 연구를 수행하였다. 다중 물체가 고유한 음향 신호를 내보내며 이동하는 상황에서 물체들의 종류와 그 위치를 추정해 내는 것이 그 목표이다. 음향 분류을 위한 전 과정으로 측정된 음향 정보로부터 주파수 영역과 시간 영역의 특징점을 추출하였으며 이를 통하여 차후 분류을 위해 필요한 훈련 데이터를 생성하였다. 기존의 음향 분류기와 제안된 분류 알고리즘의 차이는 주파수 영역 정보만을 사용한 것이 아닌 시간 영역 정보와 주파수 영역 정보를 동시에 사용한 특징점 추출 기법이다. 이를 통해 기존의 주파수 영역 정보만을 사용한 분류기보다 더 정확하고 시간 응답이 빠른 분류기를 구성하였다. 가우시안 과정을 이용한 위치 추정기법을 통해 식별되지 않은 다중 물체의 위치를 추정하였다. 최종적으로 식별되지 않는 물체의 위치 정보와 실제 물체를 연결시키는 식별 알고리즘을 제안하였다. 실험결과를 통해 분류 성능과 위치추정 및 식별 성능이 성공적으로 검증되었다.

주요어 : 다중 물체 위치 추정, 음향을 이용한 식별기법, 무선 센서 네트워크, SVM 분류기, 가우시안 프로세스 위치 추정
학 번 : 2013-20701