Statistical Filtering for Multimodal Mobility Modeling in Cyber Physical Systems

by

Arash Tabibiazar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A Cyber-Physical System integrates computations and dynamics of physical processes. It is an engineering discipline focused on technology with a strong foundation in mathematical abstractions. It shares many of these abstractions with engineering and computer science, but still requires adaptation to suit the dynamics of the physical world.

In such a dynamic system, mobility management is one of the key issues against developing a new service. For example, in the study of a new mobile network, it is necessary to simulate and evaluate a protocol before deployment in the system. Mobility models characterize mobile agent movement patterns. On the other hand, they describe the conditions of the mobile services.

The focus of this thesis is on mobility modeling in cyber-physical systems. A macroscopic model that captures the mobility of individuals (people and vehicles) can facilitate an unlimited number of applications. One fundamental and obvious example is traffic profiling. Mobility in most systems is a dynamic process and small non-linearities can lead to substantial errors in the model.

Extensive research activities on statistical inference and filtering methods for data modeling in cyber-physical systems exist. In this thesis, several methods are employed for multimodal data fusion, localization and traffic modeling. A novel energy-aware sparse signal processing method is presented to process massive sensory data.

At baseline, this research examines the application of statistical filters for mobility modeling and assessing the difficulties faced in fusing massive multi-modal sensory data. A statistical framework is developed to apply proposed methods on available measurements in cyber-physical systems. The proposed methods have employed various statistical filtering schemes (i.e., compressive sensing, particle filtering and kernel-based optimization) and applied them to multimodal data sets, acquired from intelligent transportation systems, wireless local area networks, cellular networks and air quality monitoring systems. Experimental results show the capability of these proposed methods in processing multimodal sensory data. It provides a macroscopic mobility model of mobile agents in an energy efficient way using inconsistent measurements.

Acknowledgements

I would like to express my foremost appreciation to my supervisor Professor Otman Basir for his support and guidance throughout the course of this research. I have been greatly honored to have conducted this research work under his supervision. His patience, encouragement, and thoughts have always brightened my path to the future. I am forever thankful for all I have learned from him.

I would like to thank the members of my dissertation committee, Professors Mohammed A. Quddus, Sagar Naik, Pin-Han Ho, and Matt Scott for taking the time out of their busy schedules to review my thesis and providing me with their insightful comments and suggestions.

I especially acknowledge Dr. Amir K. Khandani for his support at the beginning of my career in the University of Waterloo as a Research Associate in his Coding and Signal Transmission (CST) Laboratory.

I would like to thank Dr. Safieddin Safavi-Naeini for reviewing my Ph.D. seminar and all his supports during my career in the University of Waterloo.

I have been very fortunate to work in the Center for Pattern Analysis and Machine Intelligence (CPAMI) at the University of Waterloo. I would like to thank my fellow students in this research center.

I am thankful to the graduate studies coordinator, Wendy Boles, and Philip Regier in the Research Computing Support Group for their help during my career at the University of Waterloo.

I am grateful to my friends Dr. Daryoush Shiri, Dr. Behzad Nourani, Dr. Shahram Talakoub and all my other friends at the University of Waterloo whose company and passion for scientific research I always enjoyed.

I would like to extend my deepest gratitude to my wife, Shiva Zaboli. I am indefinitely indebted to her, and our understanding daughter, Parmis, for their sincere love, patience, and endless support, as well as my parents and sister for their unconditional love and support throughout my life. I have been fortunate to have such a great family. They all have had a constructive impact on my career.

Dedication

To my family and friends

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cyber-Physical Systems (CPS) are combinations of computation, networking, and physical processes. As an engineering discipline, CPS focused on technology, with a strong foundation in mathematical abstractions to model physical processes. Figure 1.1 shows a concept map of CPS that reveals tight interaction between information (e.g., applications, cyber security) and physical worlds (e.g., wireless sensors, control systems). This interaction requires an integration of new technologies and infrastructures (systems), and a reliable mobility model is one of the requirements in developing such new systems. This PhD thesis tackles mobility modeling as a key problem in the domain of CPS. We describe and analyze several statistical filtering methods to model the mobile agents mobility from multimodal sensory data. The proposed methods have been applied to different datasets for proof of concept.

1.1 Motivation

Mobility is one of the key characteristics in CPS. It impacts many aspects of CPS, including network capacity and routing performance. Mobility models maintain the status of mobile agents over time, including their location, velocity and acceleration changes. New communication and navigation techniques use such models frequently during their development as they can predict future user positions. In the study of a new mobile protocol, it is necessary to simulate the protocol and evaluate its performance before deployment under different communication traffic patterns.

Modeling mobility in CPS is a key problem for developing new systems and protocols. Ideally, every mobility model must capture realistic mobility patterns through a

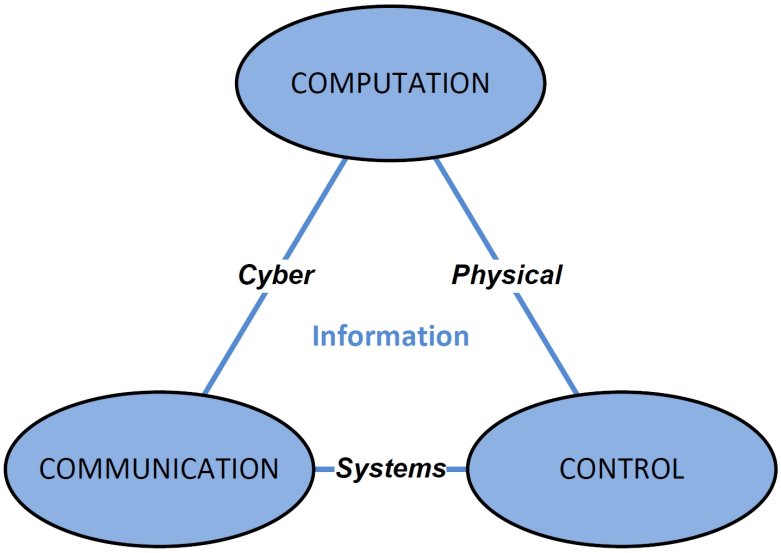

Figure 1.1: Cyber-Physical Systems

mathematically tractable approach with high flexibility in providing parametric mobility characteristics. Both analytical and simulation models can be used to describe the movement patterns or activity of a mobile user. The input to analytical mobility models requires simplified assumptions on the movement behaviors of users. These assumptions can yield performance measurements for simple cases through mathematical analysis. In contrast, simulation models consider more detailed and realistic scenarios that can derive valuable solutions for increasingly complex cases.

To extract a macroscopic model in CPS, we must process multi-modal data sources, encompassing a variety of modalities by integrating data from different sensors and methods (e.g., navigation systems, cameras, accelerometers, light sensors, Bluetooth, microphones, Wi-Fi, cell triangulation, near field communication, indoor positioning and environmental analysis). In such cases, multimodal (i.e., radio, audio and video) data fusion achieves robust and reliable results. It leverages complementing mechanisms to make the best out of the characteristics of an individual modality. Multimodal data fusion also overcomes single modality limitations and difficulties, attracting much interest in both industry and academia over the last decade.

This PhD thesis is concerned with mobility modeling in cyber physical systems. Mobility modeling is an interdisciplinary research discipline that spans interesting domains of science, including, wireless sensor networks, robotics, stochastic processes and statistical

or sparse signal processing. It revolves around a multimodal dataset that is able to capture mobility patterns in a CPS (e.g., transportation, energy, healthcare, and aerospace). In such CPS, every mobile user or element can be interpreted as a single mobile sensor. Figure 1.2 shows a variety of aspects in mobility modeling. The topic is relatively new and has received increased attention in recent years, especially when it relates to nonparametric, nonlinear filtering methods.

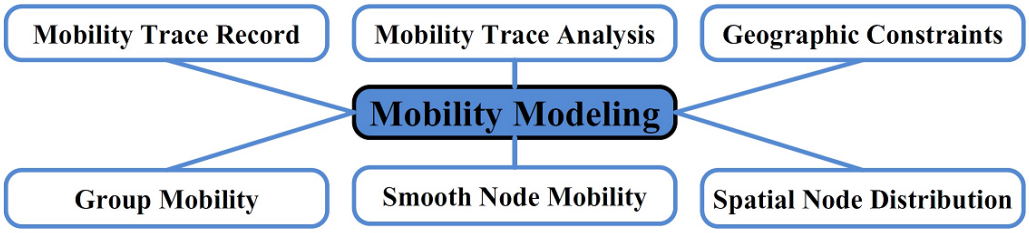

Figure 1.2: Aspects of the mobility modeling problem.

1.2 Thesis Contributions and Challenges

1.2.1 Problem Statement and Objectives

The focus of this research is on mobility modeling in CPS at the macroscopic level. Extracting a macroscopic mobility model enables the profiling of traffic conditions and hence can facilitate a wide range of applications. The thesis addresses the issue of mobility modeling through the use of statistical inference and filtering methods. Mobility in most systems is a dynamic data intensive process in which small non-linearities can lead to substantial error rates. The thesis proposes statistical filtering methods, i.e., particle filters, compressive sensing and kernel-based optimization for processing of multimodal sensory data to derive a robust mobility profile (Figure 1.3).

1.2.2 Adaptive Particle Weighting for Multimodal Data fusion

Since their introduction as optimal estimation in non-linear, and non-Gaussian state-space problems in 1993, particle filters have become a popular class of algorithms to solve such problems numerically by recursive observations. They constitute an efficient and flexible way to approximate nonlinear functions. Their efficiency and accuracy depend mainly

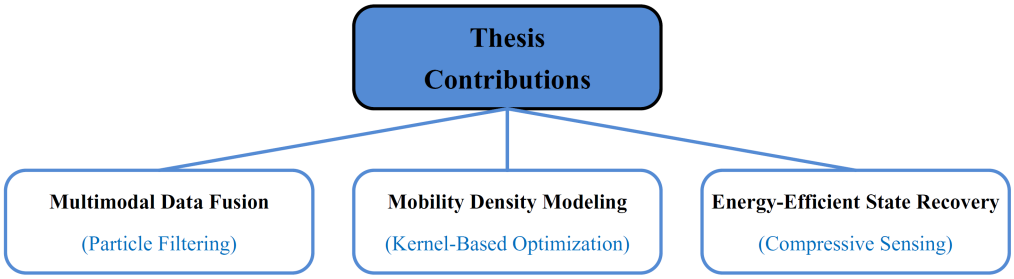

Figure 1.3: Thesis contributions to tackle mobility modeling in CPS.

on two key factors: the number of particles and the propagation function to re-allocate particles at each iteration. The thesis proposes an adaptive method to weigh the particles generated for filtering the multi-modal data. Furthermore, the thesis presents a self-adaptive particle weighted to select an appropriate number of particles and their weights when handling multimodal data.

1.2.3 Parametric Kernel Optimization for Mobility Modeling

In parametric minimum distance estimation, quantification of the degree-of-closeness between the sample data and the parametric model is a crucial step. In this thesis, the distance between a sample data and a model is formulated as a parametric function. In parametric inference, robustness is a concern, but the goal is to reduce the impact of inconsistent measurements within the model. We are interested in magnifying the small deviations from the hypothesized model to achieve greater performance. The collected position samples capture mobile user measurements (e.g., floating car data) under the assumption of Generalized Gaussian Density. The thesis proposes to use the Localized Cumulative Distribution function along with the estimated parameters to minimize the distance between the probability densities centered at mobile agent positions and the non-mobile references (anchors) as points-of-interests. The distance is minimized through quadratic programming and is subsequently used to derive the optimized weights of the kernels in the approximated probability density function of position samples. The estimated probability distribution function is used to interpolate the densities for different points-of-interests.

1.2.4 Energy-Efficient State Recovery

The mobility data and signals tend to be sparse. Recent years have seen renewed interest in the area of compressed sensing of sparse data and signals. Results in this area show that a sparse signal may be recovered from a small number of random linear measurements. To achieve this, the measurement matrix must meet certain conditions such as restricted isometry properties. The thesis proposes an energy-efficient, state recovery method for large-scale sensor networks based on the compressive sensing theory.

A sparse representation of sensory data is proposed as means to reconstruct mobility information from a few noisy measurements compared to the number of nodes. The problem is considered to be an under-determined linear system and the sparse solution provided by ℓ_1-minimization. In the case of dense signals in the time domain, it is possible to transform the signal to the frequency domain. The thesis investigates different sensing matrices to have the most energy-efficient, state recovery scheme. The performance of the proposed method is investigated with respect to power consumption and signal reconstruction error through mean-squared-error analysis. The proposed method can be adopted in large-scale sensor networks with slowly varying signals that are highly compressible. Applying this method to massive sensory data in wireless sensor networks improves resource utilization in many resource constrained applications.

Table 1.1: A list of proposed methods

Problem Aspect	Algorithm	CPS	POC Application
Multimodal Data Fusion	Particle Filters	GSM	Localization
Sparse State Estimation	Compressive Sensing	WiFi	Localization
Mobility Modeling	Kernel-Based Optimization	ITS	Density Estimation
Energy-Aware State Recovery	Compressive Sensing	WSN	Monitoring

1.3 Thesis Outline

Table 1.1 summarizes different aspects of mobility modeling problem that we tackle in this thesis. We propose solutions based on well-known modeling techniques. All solutions are evaluated in the context of relevant applications. Each solution is rigorously evaluated based on real sensor measurements such as cellular network measurement reports, floating car data in intelligent transportation system, air quality monitoring systems, wireless network measurements. Test and validation results are compared to ground truth data for performance comparison with similar methods in the literature..

The remainder of this thesis is organized as follows:

Chapter 2: Background and Literature Review

This chapter provides a brief overview of existing methods and concepts. Topics covered include probabilistic filtering, localization, mobility modeling, data fusion and compressive sensing.

Chapter 3: Multimodal Data Fusion for Localization and Mobility Tracking

This chapter describes a novel multimodal probabilistic data fusion scheme. This scheme has been used later to localize mobile users in cellular networks from inconsistent radio-visual measurements. A series of vision-based algorithms is applied to extract user locations from monocular vision. This location information is augmented with location information obtained from the cellular network. A probabilistic framework based on particle filters is developed and employed to fuse the bimodal data so as to localize the mobile user. To achieve high accuracy, an adaptive particle weighting scheme is developed based on visual information available about the surrounding of the mobile user.

Chapter 4: Mobility Estimation from Sparse Measurements

Mobility models in CPS demand high degree of accuracy in state (i.e., position, velocity, acceleration) estimation of mobile agents, especially when sparse measurements are available due to some restrictions. Accurate state estimation of mobile agents has long been considered as a challenging task in wireless networks. This chapter is concerned with the sensor state estimation problems where target sensors measure a few inconsistent signals as received-signal-strength or time-of-arrival from anchor sensors with known locations, where the target sensor location must be estimated and tracked. It is known that even in large-scale wireless sensor networks, location information is relatively sparse compared to the number of sensors. In such networks, the mobility state problem can be stated as a sparse signal recovery problem in the discrete spatial domain from a small number of linear measurements by solving an under-determined linear system. Exploitation of compressive sensing theory enables us to recover sparse signals from far fewer samples than the Nyquist sampling rate. The CS approach uses a few inconsistent measurements to find the wireless device location over a non-symmetric spatial grid. In this method, a ℓ^1-norm minimization program is used to recover the wireless user location. The system performance is measured with both synthetic and captured data indoors.

Chapter 5: Kernel-Based Optimization for Mobility Modeling

Traffic estimation and prediction services play a key role in intelligent transportation systems. The performance of such systems is heavily dependent on the availability of traffic flow information. The ability of the system to analyze the different types of floating car

data also impacts estimates. In this chapter, we propose a parametric statistical method to model the traffic flow from floating car data. The localized cumulative distribution distance between Gaussian kernels is minimized through quadratic optimization. Then, the optimized weights of the Gaussian mixtures are used to estimate the traffic density on the points-of-interest. Road geospatial coordinates are employed as optimization constraints and also as Priori knowledge on kernel means and bandwidths. A probabilistic framework is developed to extract system parameters by modeling the measurements through use of a generalized Gaussian density function. Computational complexity relaxed by linear approximation of kernel weights and its effect on performance was measured. The proposed approach is applied to real measurements of floating car data obtained from the cellular network measurement reports.

Chapter 6: Energy-Efficient State Recovery in Cyber Physical Systems

The compressive sensing theory has intensively inspired the development of the new methods and applications. Energy conservation for every node in the network and the overall power consumption are key design issues in such networks. In large-scale sensor networks, information is relatively sparse compared with the number of nodes. In such networks, the state recovery problem can be stated as a sparse signal recovery problem in the discrete spatial domain. It can be solved with a small number of linear measurements as an under-determined linear system by a ℓ_1-norm minimization program. A series of signal processing methods can be applied to recover the current state of sensory data, e.g., temperature, pressure, force, flow, humidity, position, or motion from noisy measurements. We propose an energy-efficient state recovery method for sensor networks that recovers a dense signal from sparse measurements. To obtain a sparse representation of the signal, first it is transformed to the frequency-domain, then recovered and reconstructed from a small portion of coefficients. Various random matrices driven from Bernoulli and Gaussian distributions are investigated for energy-efficient sensing in signal reconstruction. The results demonstrate more than sixty percent savings in power consumption with ten percent reconstruction and recovery error. The proposed method prolongs network lifetime with noticeable savings in deployment and maintenance costs, especially in large-scale sensor networks with slowly varying phenomena.

Chapter 7: Summary and Future Directions

This chapter discusses future research and summarizes the contributions of this thesis.

Chapter 2

Background and Literature Review

This chapter provides a brief overview of methods and concepts to be employed in this thesis. The thesis is an interdisciplinary research that spans a number of research fields, including probabilistic filters, localization, mobility models, data fusion and compressive sensing. Section 2.1 describes the probabilistic filtering methods used for data fusion. Section 2.2 is a comprehensive review on the compressive sensing theory which has been employed in energy-efficient motion state recovery and sparse signal processing. Data fusion taxonomy is reviewed in Section 2.3. In Section 2.4, we review current localization methods in wireless networks. Localization is a key problem in many research areas (e.g., robotics, wireless sensor networks, cyber-physical systems). It is also a building block in mobility models and used as a case study in. Finally, current mobility models in CPS are reviewed in Section 2.5.

2.1 Probabilistic Filtering

In this section, I briefly review the main concepts (i.e., stochastic filtering theory, Bayesian theory, and Monte Carlo theory) in probabilistic filtering theory as bases of the proposed mobility modeling. Mathematical preliminaries and background materials are also provided. Table 2.1 lists some of well-known methods on statistical filtering, inference and estimation.

Table 2.1: Statistical filtering, inference and estimation methods

Filtering	Inference	Estimation
Bayesian Filters	Bayesian Inference	Compressed Sensing
Kalman Filters	Dempster-Shafer Inference	Maximum Likelihood
Particle Filters	Fuzzy Logic	Maximum A Posterior
Stochastic Filters	Semantic Information Fusion	Least Squares
Time Series Analysis		
Hidden Markov Filter		
Gaussian Mixture Filter		
Statistical Data Analysis		

2.1.1 Bayesian Filters

British researcher Thomas Bayes originally proposed Bayesian theory in a posthumous publication in 1763. The well-known Bayes theorem describes the fundamental probability law governing the process of logical inference. However, Bayesian theory did not gain attention in its early days until its modern form was rediscovered by the French mathematician Pierre-Simon de Laplace. Bayesian inferencedevoted to applying Bayesian statistics to statistical inferencehas become one of the important branches in statistics and has been applied successfully in statistical decision, detection and estimation, pattern recognition, and machine learning. In particular, the November 19 issue of 1999 science magazine gave Bayesian research a boom in popularity in a four-page special [73]. In many scenarios, the solutions gained through Bayesian inference are viewed as optimal [125, 50].

2.1.2 Stochastic Filters

Stochastic filtering theory was established in the early 1940s as a result of Norbert Wiener and Andrey N. Kolmogorovs pioneering work and culminated in 1960 for the publication of the classic Kalman filter. Much credit should also be given to some earlier works by Bode and Shannon, and Zadeh and Ragazzini. Without any exaggeration, it is fair to say that the Kalman filters have dominated adaptive filter theory for decades in signal processing and control areas. Kalman filters have been applied in various engineering and scientific areas, including: communications, machine learning, neuroscience, economics, finance, political science and many others. Bearing in mind that the Kalman filter is limited by its assumptions, numerous nonlinear filtering methods have been proposed and developed to overcome this limitation [24]. Table 2.2 shows the development chronicle of stochastic filtering theory.

Table 2.2: Development history of stochastic filtering theory [24]

Author	Year	Method	Solution	Comment
Kolmogorov	1941	innovations	exact	linear, stationary
Wiener	1942	spectral factorization	exact	linear, stationary, infinite memory
Levinson	1947	lattice filter	approximate	linear, stationary, finite memory
Bode and Shannon	1950	innovations, whitening	exact	linear, stationary
Zadeh and Ragazzini	1950	innovations, whitening	exact	linear, non-stationary
Kalman	1960	orthogonal projection	exact	LQG, non-stationary, discrete
Kalman and Bucy	1961	recursive Riccati equation	exact	LQG, non-stationary, continuous
Stratonovich	1960	conditional Markov process	exact	nonlinear, non-stationary
Kushner	1967	PDE	exact	nonlinear, non-stationary
Zakai	1969	PDE	exact	nonlinear, non-stationary
Handschin and Mayne	1969	Monte Carlo	approximate	nonlinear, non-Gaussian, non-stationary
Bucy and Senne	1971	point-mass, Bayes	approximate	nonlinear, non-Gaussian, non-stationary
Kailath	1971	innovations	exact	linear, non-Gaussian, non-stationary
Benes	1981	Benes	exact solution of Zakai equation.	nonlinear, finite-dimensional
Daum	1986	Daum, virtual measurement	exact solution of FPK equation.	nonlinear, finite-dimensional
Gordon, Salmond and Smith	1993	bootstrap, sequential Monte Carlo	approximate	nonlinear, non-Gaussian, non-stationary
Julier and Uhlmann	1997	unscented transformation	approximate	nonlinear, non-Gaussian, derivative-free

2.1.3 State-Space Models

Dynamical systems are often modeled using a state-space framework, where the phenomenon of interest is viewed as the unknown state of the system (Figure 2.1). The state of the system is estimated based on observations- often contaminated by noise. The state-space model is often assumed to be linear:

$$x_{t+1} = F_t x_t + G_t u_t \tag{2.1}$$

$$y_t = H_t x_t + v_t \tag{2.2}$$

where $t = 0, 1, 2, ...$ is the time index, $x_t \epsilon R^m$ is the state vector, $y_t \epsilon R^k$ is the measurement vector, and $u_t \epsilon R^p$ and $v_t \epsilon R^q$ are independent noise vectors. In these equations, F_t, G_t, and H_t are known matrices. The second-order statistics of the noise vectors U_t, V_t and the initial state X_0 are known. The well-known Kalman filter recursively produces the so-called linear MMSE estimator of X_t and the linear MMSE predictor of X_{t+1} given all the data $Y_{0:t} = (Y_0, Y_1, ..., Y_t)$ available at time t. If the noise sequences and the initial state are Gaussian, the estimates are also MMSE estimates. In many problems, the noise in non-Gaussian and/or the linear model (2.2) is invalid. The performance of the Kalman filter can be disastrous in such cases, e.g., the filter may fail to track its target. The extended Kalman filter can cope with some nonlinear models by using relaxing of the state equation around the predicted state. However, this filter may still perform poorly if the relaxation is crude and/or the noise statistics are strongly non-Gaussian. This has

motivated the development of algorithms for nonlinear state-space models in which optimal MMSE estimators and predictors are produced by a nonlinear recursive filter. The state-space model is as follows:

$$x_{t+1} = f_t(x_t, u_t) \tag{2.3}$$
$$y_t = h_t(x_t, v_t) \tag{2.4}$$

where the mappings $f_t : R^m \times R^p \to R^m$ and $h_t : R^m \times R^q \to R^k$ are assumed to be known, as are the pdf's $p_{U_t}(u_t)$ and $p_{V_t}(v_t)$ for the noise vectors, and the pdf $p_{X_0}(x_0)$ of the initial state. Today particle filtering methods are widely applied to problems arising in communications, image and video processing, and statistics. The Bayesian approach requires a cost function $C(x_t, \hat{x}_t)$, the posterior pdf $p(x_t|y_{0:t})$ and $p(x_{t+1}|y_{0:t})$. For MMSE estimation, we have $C(x_t, \hat{x}_t) = \|x_t - \hat{x}_t\|^2$. The optimal estimator and predictor are given by the conditional means:

$$\hat{x}_{t|t} = \int_{R^m} x_t \, p(x_t|y_{0:t}) \, dx_t \tag{2.5}$$
$$\hat{x}_{t+1|t} = \int_{R^m} x_{t+1} \, p(x_t|y_{0:t}) \, dx_{t+1} \tag{2.6}$$

The posterior pdf can be evaluated recursively using the two-step prediction and update procedure. Due to the curse of dimensionality, evaluation of such integrals (one for each x_t!) using deterministic quadrature methods is generally computationally infeasible. The modern approach to recursive MMSE filtering is therefore to use stochastic integration methods. The particle filter, also called bootstrap filter, is a sequential Monte Carlo method which outputs a stochastic approximation to $\hat{x}_{t|t}$ and $\hat{x}_{t+1|t}$ in 2.6.

The main idea is to approximate the conditional-mean estimator and predictor of 2.8 with sample averages:

$$\tilde{x}_{t|t} = \frac{1}{N} \sum_{i=1}^{N} X_t(i) \tag{2.7}$$
$$\tilde{x}_{t+1|t} = \frac{1}{N} \sum_{i=1}^{N} X^*{}_{t+1}(i) \tag{2.8}$$

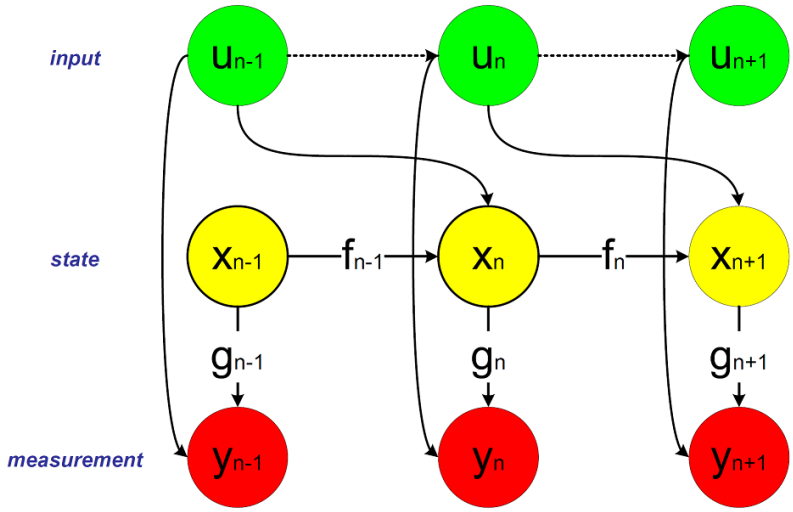

Figure 2.1: Typical State-Space Model

where the N samples $X_t(i), 1 \leq i \leq N$, are drawn iid from the posterior pdf $p(x_t|Y_{0:t})$, and similarly, the N samples $X^*_{t+1}(i), 1 \leq i \leq N$, are drawn iid from the posterior pdf $p(x_{t+1}|Y_{0:t})$. The second idea (to avoid computation of the integrals giving the posterior pdfs from which the samples are drawn) is to reuse the samples $X_t(i), 1 \leq i \leq N$, to generate $X^*_{t+1}(i), 1 \leq i \leq N$. The number of particles should be large enough so that the stochastic approximations to conditional expectations are accurate, but how large should N be? The biggest problem is the potential weak overlap of the likelihood $p(y_{t+1}|x_{t+1})$ and the prior $p(x_{t+1}|y_{0:t})$, which can cause the same samples to be reused many times, leading to an effective number of samples that is much smaller than N. This problem becomes more significant in higher dimensions. Various methods have been proposed in the statistics literature that alleviate this problem to some extent [14, 3, 98].

2.1.4 Monte-Carlo Methods

The early idea of the Monte Carlo method can be traced back to the problem of Bufion's needle when Bufion attempted in 1777 to estimate the π number. Modern formulations such the Monte Carlo methods started during the 1940s in physics and later during the 1950s in statistics. During World War II, John von Neumann, Stanislaw Ulam, Niick Metropolis and others initialized the Monte Carlo method in Los Alamos Laboratory [79]. Von Neumann also used the Monte Carlo method to calculate the elements of an inverse

matrix, in which they redefined the Russian roulette and splitting methods. In recent decades, the Monte Carlo techniques have been rediscovered independently in statistics, physics and engineering. Many new Monte Carlo methodologies (e.g., Bayesian bootstrap, hybrid Monte Carlo, quasi Monte Carlo) have been rejuvenated and developed. Roughly speaking, the Monte Carlo technique is a kind of stochastic sampling approach aiming to tackle complex systems that are analytically intractable. The power of the Monte Carlo methods is that they can attack difficult numerical integration problems. In recent years, sequential Monte Carlo approaches have attracted more and more attention to researchers from different areas with many successful applications in statistics, signal processing, machine learning, econometrics, automatic control, tracking, communications, biology and many others. One of the attractive merits of sequential Monte Carlo approaches lies in the fact that they allow on-line estimation by combining the powerful Monte Carlo sampling methods with Bayesian inference at an expense of reasonable computational cost. In particular, the sequential Monte Carlo approach has been used in parameter estimation and state estimation, for the latter of which it is sometimes called particle filter [34].

The basic concept of the particle filter is to use a number of independent random variables called particles, sampled directly from the state-space, to represent the posterior probability and update the posterior by involving the new observations; the particle system is properly located, weighted and propagated recursively according to the Bayesian rule. The formal establishment of the particle filter seems to be accurate due to Gordon, Salmond and Smith who introduced certain novel resampling techniques to the formulation. At the same time, a number of statisticians also independently rediscovered and developed the sampling-importance-resampling (SIR) idea, which was originally proposed in a non-dynamic framework. The rediscovery and renaissance of particle filters in the mid-1990s after a long dominant period was partially thanks to the ever increasing computing power. Recently, a lot of work has been done to improve the performance of particle filters. It is noted that the particle filter is not the only leaf in the Bayesian filtering tree, in the sense that Bayesian filtering can also be tackled with other techniques, such as differential geometry approach, variational method or conjugate method. Some potential future directions will be to consider combining these methods with Monte Carlo sampling techniques [24, 111]. Table 2.3 shows a list of the most popular Monte Carlo methods.

We can consider the sequential estimation of the mobility of a user within the Bayesian framework. Since the process $\mathbf{u}$ is unknown, we consider a hybrid particle that fully characterizes the target state and mode. The mobility state $\mathbf{x_k}$ can be evaluated from the conditional probability density function $p(\mathbf{x_k}|\mathbf{z_{1:k}})$ and the set of measurements $\mathbf{z_{1:k}} = \mathbf{z_1}, ..., \mathbf{z_k}$ up to time instant k via the Chapman-Kolmogorov equation:

Table 2.3: A list of popular Monte-Carlo methods [24]

Author	Method	Inference
Metropolis	MCMC	off line
Marshall	importance sampling	on/off line
N/A	rejection sampling	off line
N/A	stratified sampling	on/off line
Hastings	MCMC	off line
Geman and Geman	Gibbs sampling	off line
Handschin and Mayne	SIS	off line
Rubin	multiple imputation	off line
Rubin	SIR	on/off line
Gordon et al.	bootstrap	on line
Duane et al.	HMC	on/off line
N/A	QMC	on/off line
Chen and Schmeiser	hit-and-run MC	off line
N/A	slice sampling	off line
N/A	perfect sampling	off line

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_{R^{n_x}} p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \tag{2.9}$$

after the arrival of the measurement $\mathbf{z}_k$ at time k, the posterior state probability density function (pdf) can be updated via Bayes rule

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{2.10}$$

where $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ is a normalizing constant. This equation can be written in recursive mode as:

$$p(\mathbf{x}_k^{pf}|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k^{pf})p(\mathbf{x}_k^{pf}|\mathbf{x}_{1:k-1}^{pf})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}p(\mathbf{x}_{1:k-1}^{pf}|\mathbf{z}_{1:k-1}) \tag{2.11}$$

The analytical solution to the above equations is intractable. Hence, I utilize the sequential Monte-Carlo (particle filtering) technique as it has proven to be very suitable and powerful for dealing with nonlinear system dynamics. The MC approach relies on a sample-based construction of these probability density functions. Multiple particles (samples) of the variables of interest are generated, each one associated with a weight which characterizes the belief that the object is in this state. An estimate of the variable of

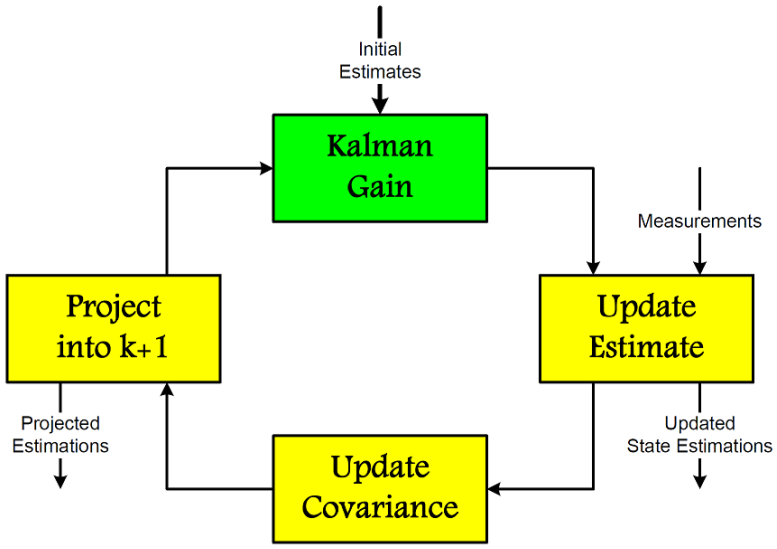

Figure 2.2: Kalman Filtering

interest is obtained by the weighted sum of particles. Two major stages can be distinguished: prediction and update. During the prediction each particle is modified according to the state model, including the addition of random noise in order to simulate the effect of the noise on the variable of interest. Then in the update stage, each particle's weight is re-evaluated based on the new sensor data. A resampling procedure deals with the elimination of particles with small weights and replicates the particles with higher weights.

A major drawback of particle filtering is that it can become prohibitively computationally expensive when a large number of particles is used. However, this computational complexity can be reduced by a procedure called Rao-Blackwellisation. Rao-Blackwellisation is a technique for improving particle filtering by analytically marginalizing some of the variables (linear, Gaussian) from the joint posterior distribution. The linear part of the system model is then estimated by a Kalman filter (Figure 2.1), an optimal estimator, while the nonlinear part is estimated by a PF. This leads to the fact that a KF is attached to each particle. In the mobility tracking problem the positions of the mobile units are estimated with a PF, while the speeds and accelerations are estimated with a KF. Since the measurement equation is highly nonlinear, the particle filter is used to approximate this distribution. After estimating the positions, these estimates are given to the KF as measurements. As a result of this marginalization process, the variance of the estimates can be reduced compared with the standard PF. Similar to the Rao-Blackwellisation ap-

proach, the mixture Kalman filtering approach proposed by Chen and Liu [22] represents the system in a linear, conditional dynamic model. In this way the problem is solved by multiple Kalman filters run in association with the MC sampling approach. A formulation of the Rao-Blackwellisation problem is given in [35].

2.2 Compressive Sensing

Conventional approaches to sampling signals or images follow Shannon's celebrated theorem: the sampling rate must be at least twice the maximum frequency present in the signal (the so-called Nyquist rate). In fact, this principle underlies nearly all signal acquisition protocols used in consumer audio and visual electronics, medical imaging devices, and radio receivers and so on. For some signals, such as images that are not naturally bandlimited, the sampling rate is dictated not by the Shannon theorem but by the desired temporal or spatial resolution. However, it is common in such systems to use an antialiasing low-pass filter to bandlimit the signal before sampling, allowing the Shannon theorem to play an implicit role. In the field of data conversion, for example, standard analog-to-digital converter technology implements the usual quantized Shannon representation: the signal is uniformly sampled at or above the Nyquist rate. CS is a novel sensing/sampling paradigm that goes against common wisdom in data acquisition. CS theory asserts that one can recover certain signals and images from far fewer samples or measurements than traditional methods would use. To make this possible, CS relies on two principles: sparsity, which pertains to the signals of interest, and incoherence, which pertains to the sensing modality.

- Sparsity expresses the idea that the information rate of a continuous time signal may be much smaller than suggested by its bandwidth, or that a discrete-time signal depends on a number of degrees of freedom which is comparably much smaller than its (finite) length. More precisely, CS exploits the fact that many natural signals are sparse or compressible because they have concise representations when expressed in the proper basis Ψ.

- Incoherence extends the duality between time and frequency and expresses the idea that objects having a sparse representation in Ψ must be spread out in the domain in which they are acquired, just as a Dirac or a spike in the time domain is spread out in the frequency domain. Put differently, incoherence says that unlike the signal of interest, the sampling/sensing waveforms have an extremely dense representation in Ψ.

The crucial observation is that one can design efficient sensing or sampling protocols to capture the useful information content embedded in a sparse signal and condense it into a small amount of data. These protocols are nonadaptive and simply require correlating the signal with a small number of fixed waveforms that are incoherent with the sparsifying basis. What is most remarkable about these sampling protocols is that they allow a sensor to very efficiently capture the information in a sparse signal without trying to comprehend that signal. Furthermore, there is a way to use numerical optimization to reconstruct the full-length signal from the small amount of collected data. In other words, CS is a very simple and efficient signal acquisition protocol which samplesin a signal independent fashionat a low rate and later uses computational power for reconstruction from what appears to be an incomplete set of measurements [19, 16].

Consider the general problem of reconstructing a vector $x \in R_N$ from linear measurements y about x of the form $y_k =< x, \phi_k >, k = 1, \ldots, K$ or $y = \Phi x$. That is, we acquire information about the unknown signal by sensing x against K vectors $\phi_k \in R^N$. We are interested in the underdetermined case $K \ll N$, where we have many fewer measurements than unknown signal values. Problems of this type arise in countless number of applications. In radiology and biomedical imaging for instance, one is typically able to collect far fewer measurements about an image of interest than the number of unknown pixels. In wideband radio frequency signal analysis, one may only be able to acquire a signal at a rate which is far lower than the Nyquist rate because of current limitations in Analog-to-Digital Converter technology. Finally, gene expression studies also provide examples of this kind. Here, one would like to infer the gene expression level of thousands of genesthat is, the dimension N of the vector x is in the thousandsfrom a low number of observations, typically in the tens. At first glance, solving the underdetermined system of equations appears hopeless, as it is easy to make up examples for which it clearly cannot be done. But suppose now that the signal x is compressible, meaning that it essentially depends on a number of degrees of freedom which is smaller than N. For instance, suppose our signal is sparse in the sense that it can be written either exactly or accurately as a superposition of a small number of vectors in some fixed basis. Then this premise radically changes the problem, making the search for solutions feasible. In fact, accurate and sometimes exact recovery is possible by solving a simple convex optimization problem[7]. In compressive sensing, we will adopt an abstract and general point of view when discussing the recovery of a vector $x \in R^N$. In practical instances, the vector x may be the coefficients of a signal $f \in R^N$ in an orthonormal basis Ψ

$$f(t) = \sum_{i=1}^{N} x_i \psi_i(t), t = 1, 2, \ldots, N \tag{2.12}$$

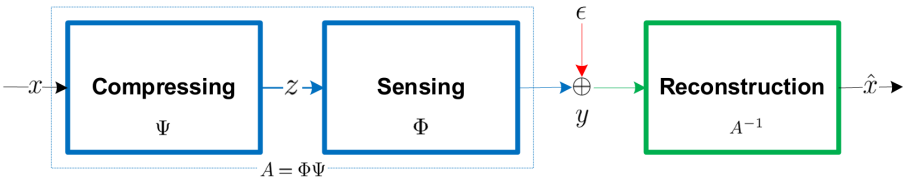

Figure 2.3: Compressive-Sensing model. Signal $\hat{x}$ is reconstructed from sparse measurements y of original signal x with added noise ϵ.

For example, we might choose to expand the signal as a superposition of spikes (the canonical basis of R^N), sinusoids, B-splines, wavelets and so on. As a side note, it is not important to restrict attention to orthogonal expansions as the theory and practice of compressive sampling accommodates other types of expansions. For example, x might be the coefficients of a digital image in a tight-frame of curvelets. To keep on using convenient matrix notations, one can write the decomposition (2.12) as $x = \Psi f$ where Ψ is the $N \times N$ matrix with the waveforms ψ_i as rows or equivalently, $x = \Psi^* f$. With this in mind, the key concept underlying the theory of compressive sampling is a kind of uncertainty relation.

If we look at data gathering approaches in wireless sensor networks, we are able to measure large amounts of data with high accuracy only by requiring the collection of a small fraction of the sensor readings. This is like a compressed sensing problem. In the past few years, the research community has been providing interesting contributions on this topic. In particular, CS is a recent compression technique that takes advantage of the inherent correlation of the input data by means of quasi-random matrices. CS was originally developed for the efficient storage and compression of digital images, which show high spatial correlation. Since the pioneering work in this area, there has been growing interest in this technique by the networking community. We have utilized this method to create a precise estimation of mobile users location from a limited number of reliable samples [113, 99].

2.2.1 Mathematical Formulation

Extensive literature on signal recovery methods from sparse or compressible signals exist. In this section, we review basic concepts in CS theory required for understanding the proposed method in this chapter. This recently emerging theory enables the reconstruction of sparse or compressible signals from a small set of non-adaptive, linear measurements in smaller sizes than the Shannon-Nyquist sampling rate [32, 18, 39].

Figure 2.3 depicts a functional model of CS as an integrated framework for compressing, sensing and reconstructing discrete time sparse or compressible signals in a known basis. Let $x \in R^n$ and $\Psi \in R^{n \times n}$ denote signal of interest and *sparsifying basis* (a.k.a. representation basis or dictionary) respectively, such that $z = \Psi x$ and $z \in R^n$ is a k-sparse vector (k non-zero elements) within n-dimensional space ($k \ll n$). CS aims to recover signal x from a set $y \in R^m$ of m linear projections of z such that:

$$y = \Phi z = \Phi \Psi x \tag{2.13}$$

where the sparsifying basis $\Psi \in R^{n \times n}$ and *sensing matrix* $\Phi \in R^{m \times n}$ (a.k.a. measurement matrix and source-channel encoder) hold specific properties. The CS theory states that a k-sparse representation z and then original signal x can be recovered from m noisy measurement vector y in polynomial time. It is the unique solution to the following linear program [16, 9].

$$\hat{z} = \arg\min_{\hat{z} \in R^n} \| \hat{z} \|_0 \quad s.t. \quad \| y - \Phi \hat{z} \|_2 \leq \eta \tag{2.14}$$

where η is a small positive number and ℓ_p-norm ($\| x \|_p$) in n-dimensional space that is defined as:

$$\| x \|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{2.15}$$

in special cases, the ℓ_1 and ℓ_0 norms of a vector are defined as the summation of absolute values and number of non-zero elements of a vector x respectively.

In some of the contributions, the sensing matrix Φ or its combination with the sparsifying matrix ($A = \Phi\Psi$) is called the measurement matrix and acts as a source encoder to express a uniform mapping between original signal and linear measurements. The only difference between projection-based source encoding in CS and conventional source coding methods in communication systems is their linearity.

CS methods are generally classified based on the type of measurement matrix (i.e., sparse or dense) measurement matrix properties to guarantee recovery process (e.g., restricted isometry property, k-neighborly polytopes, randomness extraction property, spherical section property) and reconstruction algorithms (e.g., basis pursuits, orthogonal matching pursuits, iterative thresholding).

One of the main challenges in CS is finding or designing proper sparsifying basis Ψ and sensing matrix Φ. Most of the sparsifying bases are derived from orthonormal ones, e.g., Fourier, Gabor, Curvelet or Wavelet frames with a small *coherency*, which for matrix M is defined as follows.

If M_i denotes the i^{th} column of square matrix M within n-dimensional space, then the coherence $\mu \in (0, 1)$ of this matrix is defined as:

$$\mu(M) = \max_{1 \leq i,j \leq n, i \neq j} \frac{|< M_i, M_j >|}{\parallel M_i \parallel \parallel M_j \parallel} \tag{2.16}$$

Indeed, the mutual coherence is the maximum absolute value of the cross-correlations between the columns of matrix Ψ. A dictionary is incoherent if $\mu(\Psi)$ has a small value.

2.2.2 Signal Recovery Conditions

The sensing matrices Φ can be one with deterministic or probabilistic (random) entries from well-known distributions (e.g., Gaussian, Bernoulli) that hold one of the proposed signal recovery conditions and theorems in CS. Candes and Tao have proven the main theorems of CS using either the uniform uncertainty principle (UUP) or the restricted isometry principle (RIP) [18], while Donoho has used a hypotheses called CS1 to CS3 [31, 33, 32]. Kashin and Temlyakov [59] and Zhang [124] have developed the same results with the spherical section property (SSP) that is invariant to left-multiplication by an arbitrary nonsingular matrix [117].

The $(n-m)$ dimensional subspace $V \in R^n$ has the *spherical section property* (Δ) if for any nonzero vector $v \in V$:

$$\frac{\parallel v \parallel_1}{\parallel v \parallel_2} \geq \sqrt{\frac{m}{\Delta}} \tag{2.17}$$

while m, n are positive integer numbers ($n > m > 0$). Δ is also called the *distortion* of subspace V and KGG inequality respectively [117].

One of the most prominent conditions to recover the k-sparse signal by ℓ_1-minimization algorithms is the *restricted isometry property* (RIP). It states that the k-restricted isometry

constant $\delta_k \in (0,1)$ of measurement matrix A is the smallest number, satisfied in (2.18) for all s-sparse vectors f:

$$(1-\delta_k)\| f \|_2^2 \;\leq\; \| A_s f \|_2^2 \;\leq\; (1+\delta_k)\| f \|_2^2 \tag{2.18}$$

while A_s is all possible sub-matrices consisted of s-randomly selected columns of measurement matrix A. The condition (2.18) is equivalent to having all eigenvalues of Grammian matrix $G_s = {A_s}^t A_s$ in the range of $[1-\delta_k, 1+\delta_k]$. A desirable measurement matrix in CS satisfies the RIP condition for the largest possible k values. A matrix with a small δ_k has orthonormal subsets with k or less columns. In loose terms, RIP defines the "almost-orthogonality" for every subset of measurement matrices. This range can be related to the Gershgorin circle theorem which states that eigenvalues of G_s lie in the interval $[1-\mu(k-1), 1+\mu(k-1)]$ while μ is the mutual coherency of measurement matrix A [17, 4, 29].

It may be a difficult task to verify the RIP condition for high-dimensional random matrices as there are $\binom{n}{k}$ sub-matrices to check for a k-order isometry constant for such matrices. RIP also defines a sufficient condition for holding the $\ell_0 \sim \ell_1$ equivalency [68, 18, 95, 82]. Under this condition, minimizing the ℓ_1-norm is equivalent to minimizing ℓ_0-norm. However, the former is computationally more tractable than the latter, while the latter one is very sensitive to measurement noise.

Many contributions in CS literature exist to determine the theoretic bounds as a relation between the number-of-measurements m, signal-sparsity-order k, and signal dimension n in the presence of noise, i.e., $m = \mathcal{O}(g(k,n))$. These theoretic bounds depend on the type of our sparsifying and sensing matrices and their properties such as coherency μ and isometry constant δ. For example, a CS model with a random sensing matrix of i.i.d. Gaussian entries and sparsity order $k \leq 1+1/(16\mu)$, needs at least $m = \mathcal{O}(k\log(n/k))$ measurements to reconstruct the original k-sparse signal by ℓ_1-minimization in deterministic time [95, 21, 2]. Figure 2.5 shows theoretic bounds for a random sensing matrix with Gaussian entries.

In most sensor networks with slowly varying phenomena, the state recovery sensor problem can be recast as a sparse signal recovery problem in frequency domain.

2.2.3 Signal Recovery Algorithms

According to the CS theory, the k-sparse representations of state vector $x \in R^n$ can be recovered from noisy measurements $y \in R^m$ through a linear program (4.1), where $A_{m\times n}$ is

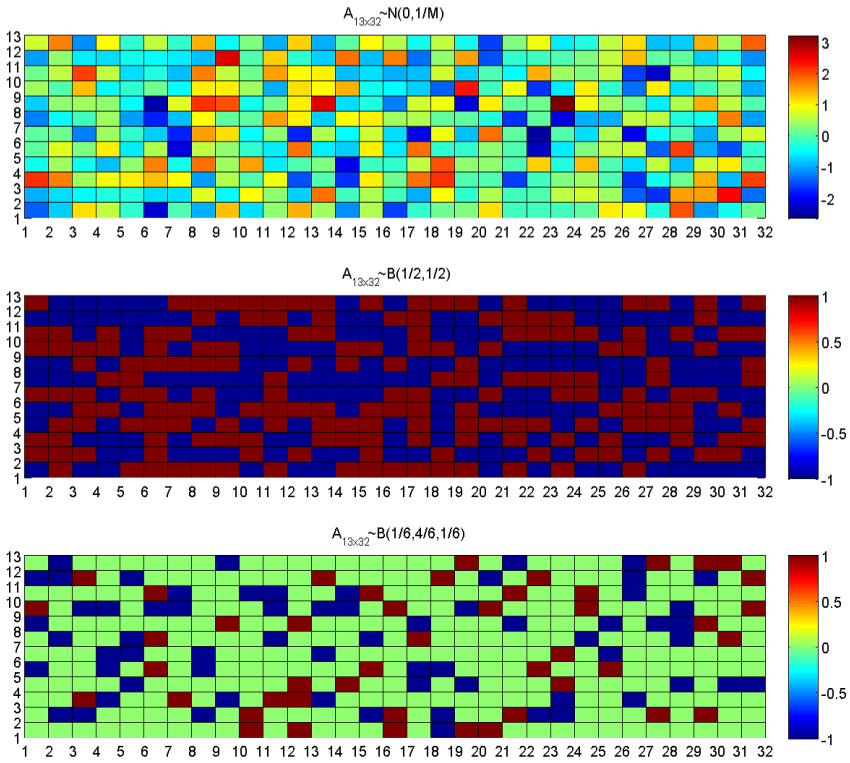

Figure 2.4: Measurement matrix $A_{13\times32}$, derived from different distribution functions. The top matrix is derived from Gaussian distribution with standard deviation $\sigma = \frac{1}{m}$. The middle one is bipolar (+1,−1) matrix derived from the Bernoulli distribution with $p = (\frac{1}{2}, \frac{1}{2})$. The bottom one is a ternary matrix, derived from mixed Bernoulli and uniform distribution with probability $p = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$ for -1, 0, and $+1$ respectively.

the measurement matrix. The results state that a k-sparse state vector x can be recovered from a small number of linear measurements y by solving an under-determined linear system:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \dots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$y = Ax + \epsilon, \quad s.t. \quad m < n \tag{2.19}$$

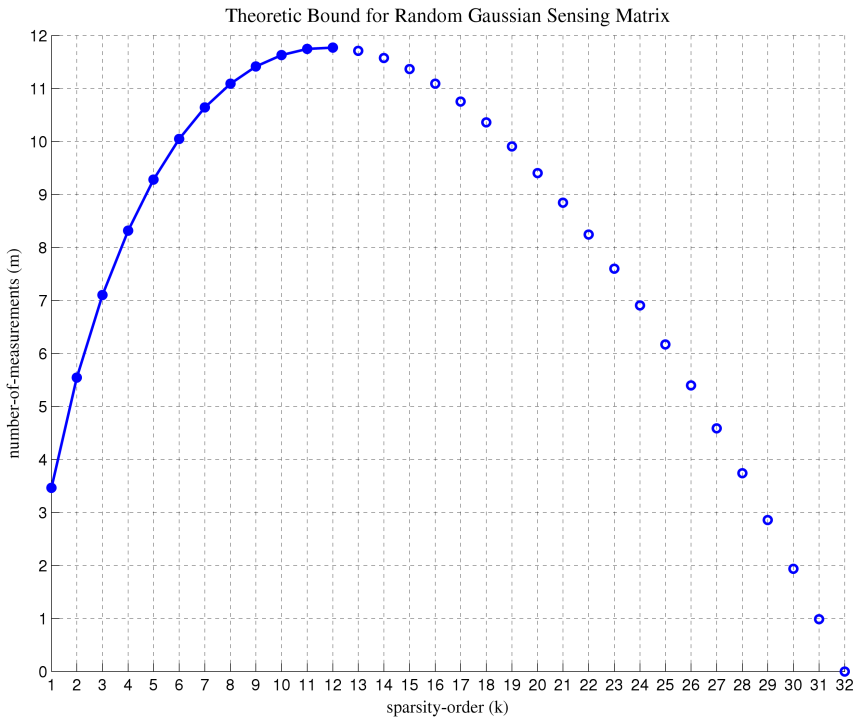

Figure 2.5: Theoretic bound for random sensing matrix with Gaussian entries. The observation-order m is plotted versus sparsity-order k . If measurement matrix A have the sparsity-order $k \leq 1 + 1/(16\mu(A))$ then to have a perfect signal reconstruction, number-of-measurements m must be in orders of magnitude $\mathcal{O}(k \log(n/k))$.

where ϵ is the measurement noise. The sparse solution of (4.1) can be found in a finite time through ℓ_1-norm minimization. However, there could be $\binom{n}{k}$ combinations of nonsingular linear systems for which to be solved. An undetermined system of linear equations has zero or infinite solutions that form an affine space [33]. Figure 2.4 shows some samples of measurement matrices that have been used in our analysis.

Although there are many algorithms presented in the literature used to find the minimums in both norm spaces, the signal processing community usually uses the ℓ_1-norm to describe a signal with just a few components [31, 66, 10]. By using ℓ_1 relaxation, problem (2.14) can be rephrased as:

$$\hat{z} = \arg\min_{\hat{z} \in R^n} \| \hat{z} \|_1 \quad s.t. \quad \| y - \Phi\hat{z} \|_2 \leq \eta \tag{2.20}$$

and k-sparse signal z can be reconstructed if sensing matrix Φ satisfies the restricted isometry property constant $\delta_{2k} < (\sqrt{2} - 1)$ [16, 43]. The recovery algorithms through ℓ_1-norm optimization are typically called "basis pursuits". The computational cost of such algorithms are usually high and can be replaced by other classes of reconstruction algorithmsknown as successive interference cancellationthat are also referred to as "matching pursuits".

$$\min_{x} \quad \| Ax - y \|_2^2 + \lambda \| x \|_1 \tag{2.21}$$

Most of the proposed methods to find the sparse solution of an underdetermined system (4.1) utilize linear programming (LP) or second-order cone programming (SOCP) to solve the problem. Our solution for problem (4.1) is based on the *l1_ls* solver (2.21) for ℓ_1-regularized least square problems [13, 23, 114].

2.3 Data Fusion

Several different terms (e.g., data fusion, sensor fusion, and information fusion) have been used to describe aspects of the fusion subject (including theories, processes, systems, frameworks, tools, and methods). Consequently, terminology confusion arises. Many definitions of data fusion have been provided through the years, most of them derived from military and remote sensing fields. In 1991, the data fusion work group of the Joint Directors of Laboratories (JDL) organized an effort to define a lexicon of terms of reference for data fusion. They define data fusion as a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single or multiple sources. Information fusion can be categorized based on several aspects. Relationships among the input data may be used to segregate information fusion into classes (e.g., cooperative, redundant and complementary data). Also, the abstraction level of the manipulated data during the fusion process (measurement, signal, feature, decision) can be used to distinguish among fusion processes. Another common classification consists in making explicit the abstraction level of the input and output of a fusion process.

- Classification based on relationship among the sources

  - Complementary: When information provided by the sources represents different portions of a broader scene, information fusion can be applied to obtain a piece of information that is more complete.
  - Redundant: If two or more independent sources provide the same piece of information, these pieces can be fused to increase the associated confidence.
  - Cooperative: Two independent sources are cooperative when the information provided by those sources is fused into new information (usually more complex than the original data) that, from the application perspective, better represents the reality.

- Classification based on levels of abstraction
  - Low-Level Fusion: Also referred to as signal (measurement) level fusion. Raw data is provided as input, combined into a new piece of data that is more accurate (reduced noise) than the individual inputs.
  - Medium-Level Fusion: Attributes or features of an entity (e.g., shape, texture, position) are fused to obtain a feature map that may be used for other tasks (e.g., segmentation or detection of an object). This type of fusion is also known as feature/attribute level fusion.
  - High-Level Fusion: Also known as symbol or decision level fusion. It takes decisions or symbolic representations as input and combines them to obtain a more confident and/or global decision.
  - Multilevel Fusion: When the fusion process encompasses data of different abstraction levelswhen both input and output of fusion can be of any level (e.g., a measurement is fused with a feature to provide a decision)multilevel fusion takes place. Typically, only the first three categories of fusion (low, medium and high level) are considered, usually with the terms pixel/measurement, feature and decision fusion.

- Classification based on input and output
  - Data In Data Out (DAI-DAO): In this class, information fusion deals with raw data and the result is also raw data, possibly more accurate or reliable.
  - Data In Feature Out (DAI-FEO): Information fusion uses raw data from sources to extract features or attributes that describe an entity.
  - Feature In Feature Out (FEI-FEO): FEI-FEO fusion works on a set of features to improve/refine a feature or extract new ones.

  - Feature In Decision Out (FEI-DEO): In this class, information fusion takes a set of features of an entity generating a symbolic representation or decision.
  - Decision In Decision Out (DEI-DEO): Decisions can be fused in order to obtain new decisions or emphasize on previous ones.

Methods, techniques and algorithms used to fuse data can be classified based on several criteria, such as the data abstraction level, purpose, parameters, type of data and mathematical foundation. The classification presented in this section is based on the methodology. According to this criterion, information fusion can be performed with different objectives such as inference, estimation, classification, feature maps, abstract sensors, aggregation and compression.

- Inference
  - Bayesian Inference
  - Dempster-Shafer Inference
  - Fuzzy Logic
  - Neural Networks
  - Abductive Reasoning
  - Semantic Information Fusion
- Estimation
  - Maximum Likelihood (ML)
  - Maximum A Posteriori (MAP)
  - Least Squares
  - Moving Average Filter
  - Kalman Filter
  - Particle Filter

We can summarize data fusion in the following items.

- Information fusion is the set of resources used to combine multiple sources so that the result is in some sense better than the individual inputs.

- Information fusion should be used to improve the performance of a task by understanding the current situation and supporting decisions.
- The techniques include filters, Bayesian and Dempster-Shafer inference, aggregation functions, interval combination functions and classification methods.
- The use of the fusion techniques should be guided by architectures and models such as the JDL model.

The provided background supports the design of fusion-based solutions for different levels of applications in a WSN, such as internal tasks (e.g., data routing) and system applications (e.g., target detection). However, there are some limitations regarding the methods and the architectures that should be considered. Depending on the model adopted, some of the listed methods might be too expensive to be executed by current sensor nodes. For example, in the Dempster-Shafer inference the combination rule has an exponential cost regarding the number of states in the frame of discernment. Thus, if two logically different states are functionally the same, from the application perspective, then they should be modeled as a single state for the sake of performance. Other methods might be improved to operate in a distributed fashion. One of the greatest challenges is to assure temporal and spatial correlation among the sources while the data is fused and disseminated at the same time. Current fusion architectures are weak in considering the peculiarities of networks because they are not network-driven. However, we understand that such architectures may be applied within specific models for networks, wherein the whole network is designed based on a global architecture for WSNs; then, the fusion task can be designed based on a fusion model that respects the requirements established by the global architecture [84].

2.4 Localization in Wireless Networks

There is extensive literature on localization methods in wireless sensor networks. Localization is a key component, especially in many context-aware applications such as pervasive medical care, smart space, wireless sensor network surveillance, mobile peer-to-peer computing, target tracking, intruder detection and environmental monitoring where position is the most essential context. A basic function in wireless positioning systems is gathering particular information around the position of a mobile device to process it later into a location estimate. This particular information could be one of the classical geolocation metrics for estimating the position (i.e., received-signal-strength, angle-of-arrival and time-of-arrival or time-difference-of-arrival). In practical environments (i.e., indoors or

outdoors), a communication source encounters environmental impacts, causing absorption, attenuation, reflection or a combination of these effects on a communication medium. The accurate localization of objects and people has long been considered an important building block in wireless systems. In most research fields, short-range radio signal measurements (i.e., WiFi, Bluetooth, ultrasound or infrared) are used as input data for localization technique [87, 12, 36, 70].

Due to hardware constraints and environmental effects, noisy measurements are inevitable in all range-based (i.e., time of signal arrival, the angle of signal arrival, the time difference of arriving signals and the received signal strength) methods. As a result, many signal noise-cancellation and estimation techniques have been proposed to infer true states from multiple-sensor measurements. Localization within a Bayesian probabilistic framework has achieved good performance in robotics and recently been successfully applied to sensor networks. Probabilistic inference is typically made by observing multiple sets of noisy evidences. Particle filtering uses a probabilistic transition and observation model model to estimate the user state in every iterations. The key in particle filtering is the representation of the posterior distribution by a set of weighted samples (or particles) in which samples represent a probability distribution function. Range-based localization introduces non-Gaussian components in the measurements due to the nonlinearity in wireless fading channel. The mobility introduces extra difficulty in localization, but reveals certain information that we can benefit from this information in both filtering and smoothing methods [60, 97].

Range-based localization methods are the most simple and inexpensive way to estimate the use location in wireless network but suffer from noisy and inaccurate measurements or delays due to fading in wireless channel. Most of these localization methods first estimate the range based on an empirical data model and then infer position coordinates using a filtering method (e.g., Kalman filter) [70, 81].

All of these factors make localization systems a key technology for the development and operation of WSN. At present, Mobile Station location information is important for any service provider. For a known location, various services can be provided. Therefore, location information plays an important role for the next generation of mobile systems. The accuracy of location information is important from the users point of view. Some services require higher accuracy than other services and therefore choosing the positioning method for a specific service is important. One viewpoint for examining positioning systems is to consider where the position measurements are made and how the position information is used. Three broad classifications are made: self-positioning (mobile-based positioning), remote positioning and hybrid positioning.

- Self-positioning: positioning receiver makes the necessary signal measurements from geographically distributed transmitters and uses these measurements to determine actual position.
- Remote positioning: receivers measure a signal originating from the positioned station. These measurements are transferred to a central site and they are combined to give an estimate of mobile location.
- Hybrid positioning: it is possible to send the location measurement data from a self-positioning receiver to a remote site and there is realized calculation process or vice versa.

The type of measurement depends on the positioning method (i.e., signal strength, time of arrival). In range-based localization with received-signal-strength, the propagation path-loss is measured and converted to the distanc between transmitter and receiver. For two-dimensional positioning every measurement corresponds to a circle centered at the transmitter where the receiver lies in this circle (Figure 2.6). In the absence of measurement error, the position is given by the intersection of circles from at least three transmitters in order to resolve ambiguities arising from crossing multiple lines of position (circles). Signal propagation conditions and environment have a significant impact on localization accuracy. The changes in the transmission path are characterized by various influences, e.g., shadowing, multipath fading and path loss. The shadowing is characterized by long-term fading due to terrain contour between transmitter and receiver. It represents a slow variation in the mean envelope over a distance. The multipath fading is caused by NLOS propagation. The NLOS propagation always exists in cities or other built-up environments. The multipath fading or short-term fading is characterized by the fast variation of received signal strength over a short distance on the order of a few wavelengths or short time durations. The path loss indicates signal power attenuation with the increasing distance between the transmitter and the receiver. These influences cause deviation of signal strength measurement and degrade the location accuracy. The variations of received signal strength can acquire values up to 40dB in some environments[36].

A CPS consist of a collection of wireless networked, low-power sensor devices that include nodes that integrate an embedded microprocessor, radio and a limited amount of storage. Localization in this context inherently poses unique challenges because of imperfect sensor characteristics, limited computational capability and inaccurate distributed sensor fusion. Contemporary sensor technologies that are used in localization mainly fall into the categories of ultrasonic, infrared, radio and laser types. For practical applications, the resolution requirements are also quite different. A vital problem in signal strength

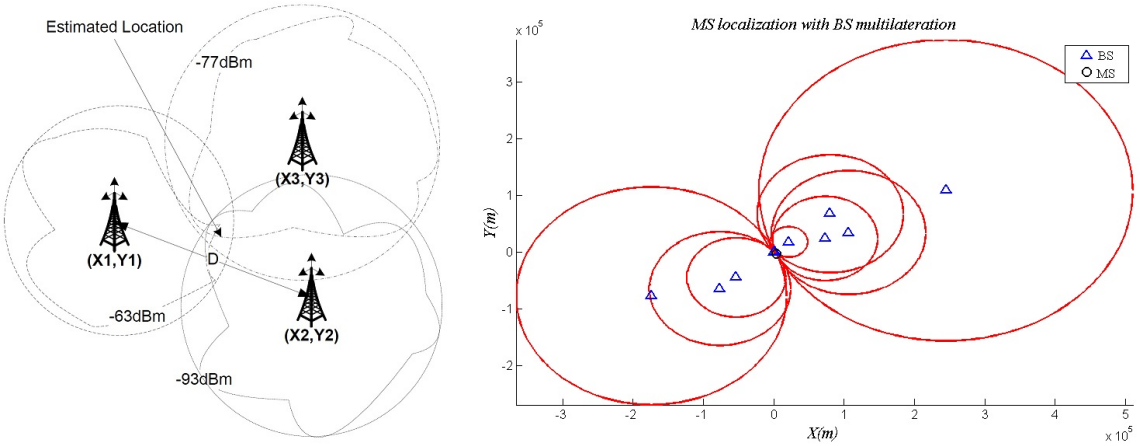

Figure 2.6: Mltilateration in a wireless networks using inconsistent measurements. It has not yet overcome the challenges of poor ranging measurements in dynamic and noisy environments.

based localization methods is how to deal with varying power levels. The ability of dynamically changing the transmit power at runtime is promising in current sensor networks to save energy and reduce interference with each other. Due to the dynamically changing power levels, most previous static-power-level-based localization algorithms no longer work. Radio transmit-power adjustments can provide rich information about the signal strength, which can further provide evidence of node positions using a filtering method. There is a probabilistic world during the process of data propagation due to the nature of wireless environments, such as failure-prone wireless links, densely deployed sensors and noisy measurements [51].

In many research fields other than wireless networks, localization is a part of another problem. For example, in robotics, it is integrated in Simultaneous Localization and Mapping (SLAM) problem. Conventionally, most work on SLAM has focused predominantly on active range sensors such as lasers for data association. Laser sensors have high-depth resolution that provides accurate measurements but are expensive, heavy, and highly power consumed. They suffer from the perceptual aliasing problem. On the other hand, cost effective cameras are becoming more and more popular in the SLAM community because of reliable vision tracking algorithms [42, 5, 37]. Moreover, cameras with unique features might lead to more reliable data association than laser scanners. Most vision based approaches use binocular or trinocular stereos that are similar to range finding devices [45] in the sense that they concentrate on solving the pose estimation problem from the discrep-

ancy image features. Stereo geometry requires the corresponding pattern to be visible in both cameras, therefore, cameras need to undergo synchronous calibration and be placed relatively close together. However, the accuracy of pose estimation depends on the spacing between the cameras and the synchronous calibration. As small camera spacing results in low depth resolution and asynchronous calibration degrades accuracy. Conversely, single camera allows that the camera placement to be more optimal with respect to occlusion and accuracy. In a reduced dimension, some researchers have pointed single cameras toward the floor or ceiling [58]. Folkesson et al. similarly exploits the visual and geometric salience of the walls between those two extremes [41]. The use of general monocular vision has recently drawn due to several problems related to using a single camera. First, At least two observations of the same landmark - from two sufficiently spaced locations - are necessary to initialize mobility state vector. Second, the landmark observation consistencies from those two locations depend on the stability of the tracking algorithm as well as the camera location. Third, the high uncertainty in initialization may lead to the divergence of the SLAM algorithm and imprecise data association. However, these constraints can be manageable with proper initialization algorithms. Recent approaches in SLAM can be divided into bottom-up and top-down approaches. They are natural inheritances of structure-from-motion (SFM) approaches of the vision community and Kalman Filter (KF) approaches of the control community. Bottom-up approaches use a batch nonlinear filter adjustment or a batch EKF procedure to initialize the landmarks from past observations. Several bottom-up approaches use a large database of feature descriptors to localize the robot [67]. Alternatively, top-down approaches [27] acknowledge the fact that feature depth uncertainty during initialization is not well-modeled by a standard Gaussian distribution in Euclidean space. In delayed initialization style, observations of features are pruned using observation evidences until the most likely single Gaussian samples remained. They are added to the filter to update the camera pose estimate. In non-delayed initialization, multi-hypotheses distributions are explicitly imposed on the state. However, the convergence of the filter when updating a multi-Gaussian feature and real time performance is not proved [93]. It can also impose indoor environmental constraints to improve the map [112].

A solution for visual navigation with reference to a pre-recorded image sequence was presented by Yoshio Matsumoto et al. [75] but the method did not disclose the pose of the robot at each frame. After a human guided phase, the robot builds a map of the environment and uses it to localize itself in autonomous navigation. This system relies both on vision and odometry. Recently, the idea of using a visual memory has been proposed in which the path is modeled as a set of key frames and the robot goes from one frame to the next by visual serving. This approach avoids building a 3D reconstruction and therefore

is not able to provide a 3D localization of the robot.

Several methods for localization using vision have been proposed by Se et al. [102]. Real-time SLAM using only monocular vision - in small environments with less than 100 landmarks - has been achieved by Davison [27]. It is also possible to compute ego motion by using only visual data as done by Nister et al. [86, 85]. In this case, maintaining a large map is not required but the localization accuracy decreases as the distance traveled increases. Moreover, in two successive navigation experiments, the robot may not reuse the same landmarks and the resulting trajectory may be different. Another approach for achieving robot navigation according to a human-guided experience consists in representing the trajectory as a set of key images. Then the robot has to travel from one key frame to the next. With this approach, the robot will go through a number of well-defined positions and the trajectory is repeatable from one navigation experiment to the next. The first possibility to do that is to go from one key frame to the next by visual serving as done by Matsumoto et al. [75]. Another possibility was presented by Goedeme et al. [44]. In this method, a relative localization of the robot with reference to the key frame is computed by features matching using a wide baseline techniques. After that, a displacement vector is computed from next key frame. Some approaches begin with building a map of the environment offline. After that, robot is able to localize itself with this map. This is not always the case in SLAM approaches because map building and localization must be done simultaneously in real-time. The map can be built by using different techniques and sensors. Vacchetti et al. [115] use a CAD model - provided by the user - and develop an efficient localization system. Cobzas et al. [25] use a rotating camera along with a laser range finder to build a set of panoramic images enhanced with 3D information. With this map, a single 2D image is enough to localize the camera. Kidono et al. [61] build a 3D reconstruction from a stereo camera and an odometer under the assumption that the ground is planar. They use this map later to localize the robot in real-time.

2.5 Mobility Models

We usually study the mobility models in dynamic environments like CPS or mobile ad hoc networks (MANET). MANET is a collection of wireless nodes communicating with each other in the absence of any infrastructure. Due to the availability of small and inexpensive wireless communication devices, the MANET research field has attracted a lot of attention from academia and industry in the recent years. In the near future, MANETs could potentially be used in various applications such as mobile classrooms, disaster relief applications and battlefield communications. To thoroughly and systematically study a new MANET

protocol, it is important to simulate this protocol and evaluate its performance. Protocol simulation has several key parameters, including mobility model and communicating traffic pattern.

2.5.1 Mobility Models in Ad-hoc Networks

The mobility model is designed to describe the movement pattern of mobile users, and how their location, velocity and acceleration change over the time. Since mobility patterns may play a significant role in determining the protocol performance it is desirable for mobility models to emulate the movement pattern of targeted real life applications in a reasonable way. Otherwise, the conclusions drawn from the simulation studies may be misleading. Thus, when evaluating MANET protocols, it is necessary to choose a proper underlying mobility model. For example, the nodes in a random waypoint model behave quite differently as compared to nodes moving in groups. It is not appropriate to evaluate the applications where nodes tend to move together using a random waypoint model. Therefore, a real need for developing a deeper understanding of mobility models and their impact on protocol performance exists. One intuitive method to create realistic mobility patterns would be to construct trace-based mobility models, in which accurate information about the mobility traces of users could be provided. However, since MANET have not been implemented and deployed on a wide scale, obtaining real mobility traces becomes a major challenge. Therefore, various researchers proposed different kinds of mobility models, attempting to capture various characteristics of mobility and represent mobility in a somewhat 'realistic' fashion. Much of the current research has focused on the synthetic mobility models that are not trace-driven. In the previous studies on mobility patterns in wireless cellular networks researchers mainly focused on the movement of users relative to a particular area (like cell) at a macroscopic level, such as cell change rate, handover traffic and blocking probability. However, to model and analyze mobility models in MANET, we are more interested in the movement of individual nodes at the microscopic-level as these factors directly determine when the links are formed and broken[15, 1].

One frequently used mobility model in MANET simulations is the random waypoint model, in which nodes move independently to a randomly chosen destination with a randomly selected velocity. The simplicity of a random waypoint model may have been one reason for its widespread use in simulations. However, MANET may be used in different applications where complex mobility patterns exist. Hence, recent research has started to focus on alternative mobility models with different mobility characteristics. In these models, the movement of a node is more or less restricted by its history or other nodes in the neighborhood or the environment.

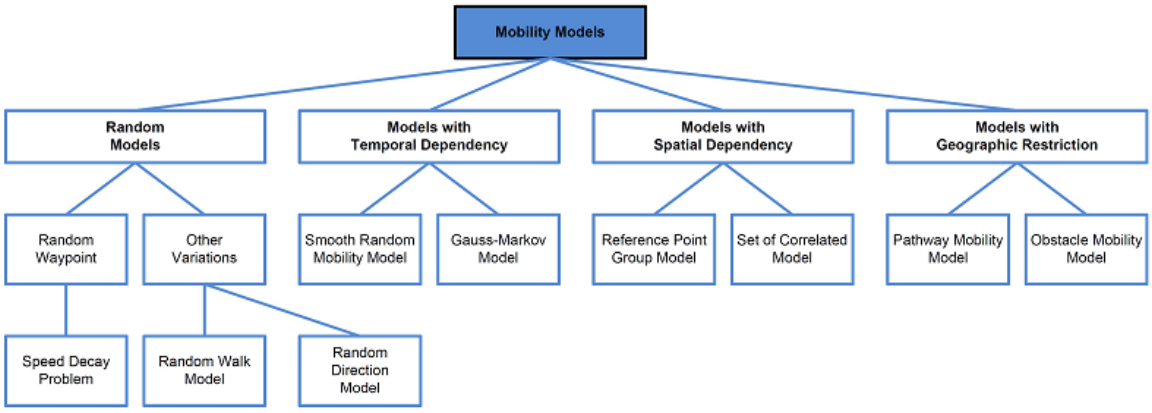

Figure 2.7: The categories of mobility models in ad-hoc networks [1].

In Figure 2.7, a categorization for various mobility models into several classes based on their specific mobility characteristics is provided. For some mobility models, the movement of a mobile node is likely to be affected by its movement history. This type of mobility model is referred to as a mobility model with temporal dependency. In some mobility scenarios, the mobile nodes tend to travel in a correlated manner. We refer to such models as mobility models with spatial dependency. Another class is the mobility model with geographic restriction, where the movement of nodes is bound by streets, freeways or obstacles. In general, the mobility models can be classified according to the different kind of dependencies and restrictions that are considered.

1. Random based: There are neither dependencies nor any other restriction modeled.
2. Temporal dependencies: The actual movement of a node is influenced by the movement of the past.
3. Spatial dependencies: The movement of a node is influenced by the nodes around it (group mobility).
4. Geographic restrictions: The area in which the node is allowed to move is restricted.
5. Hybrid characteristics: A combination of temporal dependencies, spatial dependencies and geographic restrictions is realized.

On the other hand, mobility models have been long recognized as one of the fundamental components that impacts the performance of wireless ad hoc networks. A wide

variety of mobility models are available in the research community. Among all mobility models, the popularity of random mobility models (e.g., random walk, random direction and random waypoint) is rooted in its simplicity and mathematical tractability. A number of important properties for these models have been studied, such as the stationary nodal distribution, the hitting and meeting times, and the meeting duration. These quantities in turn enable the routing protocol analysis to produce performance bounds. However, random mobility models are based on over-simplified assumptions, and as has been shown recently, the resulting mobility characteristics are very different from real-life scenarios. Hence, it is debatable whether the findings under these models will directly translate into performance in real-world implementations of MANET. More recently, an array of synthetic mobility models are proposed to improve the realism of the simple random mobility models. More complex rules are introduced to make the nodes follow popularity distribution when selecting the next destination, stay on designated paths for movements or move as a group. These rules enrich the scenarios covered by the synthetic mobility models, but at the same time make theoretical treatment of these models difficult. In addition, most synthetic mobility models are still limited to iid models, and the mobility decisions are also independent of the current location of nodes and time of simulation. A different approach to mobility modeling is by empirical mobility trace collection. Along this line, researchers have exploited existing wireless network infrastructure, such as wireless LANs or cellular phone networks, to track user mobility by monitoring their locations. Such traces can be replayed as input mobility patterns for simulations of network protocols. More recently, DTN-specific testbeds were introduced aimed at collecting encounter events between mobile nodes instead of the mobility patterns. Yet, the size of the traces and the environments in which the experiments are performed cannot be adjusted at will by the researchers. To improve the flexibility of traces, the approach of trace-based mobility models have also been proposed. These models discover the underlying mobility rules that lead to the observed properties in the traces. Statistical analysis is then used to determine proper parameters of the model to match it with the particular trace [56, 62, 104].

In the literature, vehicular mobility models are usually classified as either macroscopic or microscopic. The macroscopic description models gross quantities of interest, such as vehicular density or mean-velocity and treat vehicular traffic according to fluid dynamics. The microscopic description considers each vehicle as a distinct entity, modeling its behavior in a more precise, but computationally more expensive way. Yet, a micro-macro approach may be seen more as a broad classification schema than a formal description of the models' functionalities in each class. A more precise way that may be suggested for looking at mobility models consists in identifying functional blocks: motion constraints, traffic generator, time and external influences. On the other hand, motion constraints de-

scribe the relative degree of freedom of each vehicle. Macroscopically, motion constraints are streets or buildings, but microscopically, constraints are modeled by neighboring cars, pedestrians, or by diversities either due to the type of car or due to the driver's habits. On the other hand, the traffic generator defines different kinds of cars and deals with their interactions according to the environment under study. Macroscopically, it models traffic densities, speeds and flows, while microscopically it deals with properties like the inter-distance between cars, acceleration, braking and overtaking. Another important aspect of realistic motion modeling is time, which can be seen as the third functional block that describes different mobility configurations for a specific time of the day or day of the week. Finally, we also have to add a fourth fundamental block, the External Influence, modeling the impact of a communication protocol or any other source of information on the motion patterns. According to the concept map in Figure 2.7, mobility models intended to generate realistic vehicular motion patterns should include the following building blocks:

- Accurate and Realistic Topological Maps: Street topologies should manage different densities of intersections, contain multiple lanes, different categories of streets and their associated speed limitations.
- Obstacles: Obstacles should be understood in a wide sense, as both constraints to car mobility and hurdles to wireless communication.
- Attraction/Repulsion Points: Initial and final destinations of road trips are not random. Most of the time, drivers are moving to similar final destinations, called attraction points (e.g., offices), or from similar initial locations, called repulsion points (e.g., homes), a feature that creates bottlenecks.
- Vehicles Characteristics: Each category of vehicle has its own characteristics, which has an impact on set of traffic parameters. For example, macroscopically speaking, some urban streets and highways are forbidden to trucks depending on the time of the day. Microscopically speaking, acceleration, deceleration and speed capabilities of a car or a truck are different. Accounting for these characteristics alters the traffic generator engine when modeling realistic vehicular motions.
- Trip Motion: A trip is macroscopically seen as a set of source and destination points in the urban area. Different drivers may have diverse interests, which affect its trip selection.
- Path Motion: A path is macroscopically seen as the set of road segments taken by a car on its trip between an initial and destination point. As it may also be observed

in real life, drivers do not randomly choose the next heading when reaching an intersection, as it is currently the case in most vehicular networking traffic simulations. Instead, they choose their paths according to a set of constraints such as speed limitations, time of the day, road congestion, distance and even driver's personal habits.

- Smooth Deceleration and Acceleration: As vehicles do not abruptly break and accelerate, models for decelerations and accelerations should consequently be considered.
- Human Driving Patterns: Drivers interact with their environments, not only with respect to static obstacles, but also to dynamic obstacles, such as neighboring cars and pedestrians. Accordingly, the mobility model should control the mutual interactions between vehicles, such as overtaking, traffic jam and preferred paths.
- Intersection Management: It corresponds to the process of controlling an intersection and may either be modeled as a static obstacle (stop signs), a conditional obstacle (yield sign), or a time-dependent obstacle (traffic lights). It is a key part in this framework that however only influences the Motion Constraint block, as the Traffic Generator block cannot see the difference between a stop sign or high-density traffic. Both are interpreted as a motion constraint.
- Time Patterns: Traffic density is not identical during the day. A heterogeneous traffic density is always observed at peak times, such as at rush hours or during special events. This block influences the Motion Constrains and the Traffic Generator blocks, as it may alter the trip or path computation and also the attraction/repulsion points.
- External Influence: Some motion patterns cannot be proactively configured by vehicular mobility models as they are externally influenced. This category models the impact of accidents, temporary road works or real-time knowledge of the traffic status on the motion constraints and the traffic generator blocks. Communication systems are the primary source of information about external influence [103].

All building blocks described here should be followed by designers of mobility models specific to vehicular motions. The more building blocks a vehicular mobility model includes, the more realistic it is.

2.5.2 Traffic Flow Models

This section reviews the traffic flow models in intelligent transportation systems as a subcontext in traffic flow modeling. Different traffic models have been developed over nearly

eight decades that can be broadly categorized into microscopic, mesoscopic and macroscopic (Figure 2.8) based on their description that details various aspects of traffic flow operations. Microscopic models seek to represent traffic flow in terms of individual driver time-space behavior processes, while mesoscopic models describe traffic flow dynamics without explicitly distinguishing time-space in microscopic models. Our focus in this thesis is a macroscopic traffic flow model for mobility modeling. In the macroscopic level, the global behavior of traffic stream is modeled in terms of quantities for a group of vehicles such as volume, density and velocity [94, 54].

Bruce Greenshields studies on traffic flow in the early 1930s resulted in the first speed-flow curves, which published in 1935. For the first time, he carried out some tests to measure traffic flow, density and speed using photographic measurement.

A first-order macroscopic model of traffic flow is proposed by Lighthill, Whitham and Richard (LWR) for the first time [69] and later second-order models are introduced [96, 120]. They describe the aggregated traffic state (e.g., density or speed) without considering its basic entities (i.e., driver behavior). This level of aggregation reduces the computational and calibration costs. It is proven that the macroscopic models are more robust than the microscopic or mesoscopic models in describing the traffic flow dynamics based on the averaged or aggregated information of individual driver behaviors.

The macroscopic model typically employs the n^{th}-order continuum equation (2.22) to describe the dynamics of traffic flow. It implies that the number of leaving vehicles is equal to entering ones plus the number of current vehicles in the same segment. The first-order continuum model is also known as the LWR model. It is the most popular practical model due to its simplicity and performance.

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0 \tag{2.22}$$

Another equation is the basic traffic flow equation (2.23) that describes the volume of traffic in terms of density and average speed:

$$q = k.u \tag{2.23}$$

In both equations, q stands for volume, k for density, u for speed, x for space and t for time respectively.

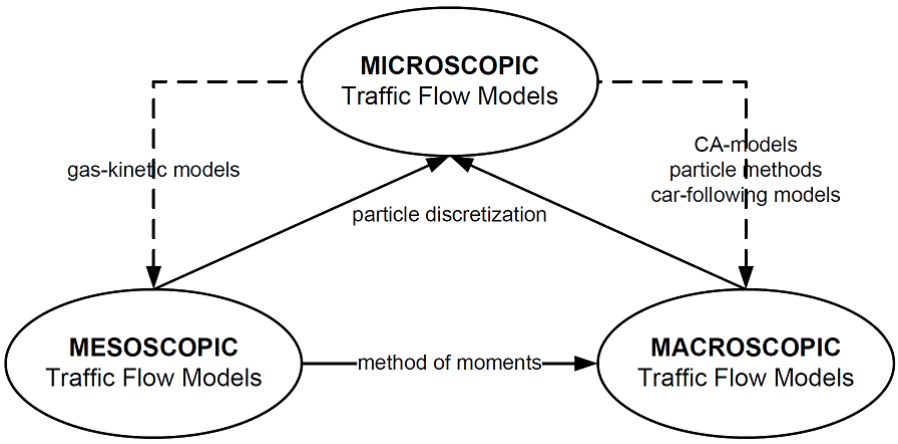

Figure 2.8: Traffic flow models: microscopic, mesoscopic, and macroscopic [54].

Other than mathematical discontinuity of the LWR models, their shortcomings (e.g., stop-start characterization and non-stationary dynamics) have been addressed in several works [65]. Most of deficiencies in the LWR models can be tackled by utilizing the higher-order continuum equations [92, 120]. There were many questions on the performance of higher-order continuum equations compared to LWR models, especially in congestion situations [80, 26, 88, 54]. The studies of Papageorgiou et al. on traffic models with real field data under different traffic conditions are shown outperforming the higher-order continuum models over the first-order ones [90, 89]. Later, Liu et. al. [72] and Zhang [123] improved the higher-order continuum models that were consistent with LWR theory and then provided the numerical results.

The macroscopic traffic flow has been used in many simulation models (e.g., FREFLO, NETFLO, METANET, MASTER and TRANSYT) to predict the traffic condition by aggregating the traffic data. Due to a lower level of accuracy, they are usually appropriate for static planning in large areas [94]. A complete survey of vehicular traffic modeling methods can be found in [54].

One can find many other approaches, i.e., statistical, linear regression, Markovian, cellular automaton, queuing theory, fuzzy, neural and pattern recognition models on flow models all with their pros and cons. In all of the proposed methods, precise prediction of traffic flow is very dependent on the quality of the traffic data and processing technique used [55].

In probabilistic filtering, the dynamics of the mobile agent can be described by the equation:

$$\mathbf{x_k} = \mathbf{f}(\mathbf{x_{k-1}}, \mathbf{m_k}, \mathbf{u_k}, \mathbf{w_k}) \tag{2.24}$$

where $\mathbf{x_k} \in R^{n_x}$ is the system base state, $\mathbf{u_k} \in R^{n_u}$ specifies the command process, and $\mathbf{w_k} \in R^{n_w}$ is the state noise, with $k \in N$ being the discrete time and N is the set of natural numbers. The modal (discrete) state $\mathbf{m_k}$ characterizing the different system modes (regimes), can take values over a finite set M , i.e., $\mathbf{m_k} \in M$. The measurement equation has the form of:

$$\mathbf{z_k} = \mathbf{h}(\mathbf{x_k}, \mathbf{v_k}) \tag{2.25}$$

where $\mathbf{z_k} \in R^{n_z}$ is the observation and $\mathbf{v_k} \in R^{n_v}$ is the measurement noise. Functions $f(.)$ and $h(.)$ are nonlinear in general. It is assumed that the observations are taken at discrete time points T_k, with a discretized time step T. The acceleration u_k of the mobile unit is usually highly correlated, but sometimes it undergoes rapid changes caused by various reasons such as traffic lights and road turns. The motion of the mobile user can be modeled as a dynamic system driven by a command $\mathbf{u_k} = (u_{x,k}, u_{y,k})$ and a correlated random acceleration $\mathbf{r_k} = (r_{x,k}, r_{y,k})$ at time k, i.e., the total acceleration is $\mathbf{a_k} = \mathbf{u_k} + \mathbf{r_k}$.

The state of mobile agent at time instant k is defined by the vector $\mathbf{x_k} = (x_k, \dot{x}_k, \ddot{x}_k)$ where x_k and y_k specify the position, $\dot{x}_k$ and $\dot{y}_k$ specify the speed, and $\ddot{x}_k$ and $\ddot{y}_k$ specify the acceleration in the x and y directions in a two-dimensional space. The motion of the mobility user can be described by the equation:

$$\mathbf{x_k} = \mathbf{A}(\mathbf{T}, \alpha)\mathbf{x_{k-1}} + \mathbf{B_u}(\mathbf{T})\mathbf{u_k} + \mathbf{B_w}(\mathbf{T})\mathbf{w_k} \tag{2.26}$$

$$\begin{pmatrix} x_k^{(j)} \\ \dot{x}_k^{(j)} \\ \ddot{x}_k^{(j)} \\ y_k^{(j)} \\ \dot{y}_k^{(j)} \\ \ddot{y}_k^{(j)} \end{pmatrix} = \begin{pmatrix} 1 & T & T^2/2 & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & T^2/2 \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & \alpha \end{pmatrix} \times \begin{pmatrix} x_{k-1}^{(j)} \\ \dot{x}_{k-1}^{(j)} \\ \ddot{x}_{k-1}^{(j)} \\ y_{k-1}^{(j)} \\ \dot{y}_{k-1}^{(j)} \\ \ddot{y}_{k-1}^{(j)} \end{pmatrix} + \begin{pmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & 0 \\ 0 & T^2/2 \\ 0 & T \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} u_{x,k}^{(j)} \\ u_{y,k}^{(j)} \end{pmatrix} + \begin{pmatrix} T^2/2 & 0 \\ T & 0 \\ 1 & 0 \\ 0 & T^2/2 \\ 0 & T \\ 0 & 1 \end{pmatrix} \times N(0, \sigma_{w,k}^2 I_{2\times 1})$$

The unknown command processes $\mathbf{u_{x,k}}$ and $\mathbf{u_{y,k}}$ are modeled as a first-order Markov chain that takes values from a set of acceleration levels M_x and M_y.

2.6 Chapter Summary

In this chapter, we reviewed the basic concepts in probabilistic filtering, localization, traffic models, data fusion and compressive sensing. A comprehensive review of existing compressive sensing methods and conditions for signal recovery from sparse measurements were reviewed.

Chapter 3

Multimodal Data Fusion for Localization and Mobility Tracking

This chapter describes a novel multimodal data fusion scheme that is used in localization of mobile users in cellular networks. A series of vision-based algorithms is applied to derive a preliminary mobile user position from monocular vision to augment it with estimated location in a cellular network. A probabilistic framework has been developed based on particle filters to integrate bimodal data (video and radio) and recover mobility pattern of mobile users in cellular networks from inconsistent signal-strength measurements. An adaptive particle weighting scheme based on the modal confidence coefficient is developed. This approach can be easily implemented to utilize online visual databases. The evaluation show that, the proposed method has higher accuracy than conventional positioning methods in wireless networks even in indoors when accurate navigation signals (e.g., GPS) are not available. Figure 3.1 is a sketch of the problem that we tackle in this chapter.

The most common positioning sensor for outdoors is the GPS device that allows highly accurate localization. Such accuracy is impacted by a number of factors, including satellite positions, noise in the radio signal, atmospheric conditions, and natural barriers in the signal path. Unfortunately, in dense urban areas or indoors, buildings can mask received signals and prevent accurate localization significantly. It is necessary to use other data sources (i.e., video, audio signals) to solve localization problem especially where the odometry is difficult. Augmenting the received signal strength information with extracted positions from visual signals may lead to finding user positions precisely.

One desired application of the proposed method can be delivering the location based services to mobile users in cellular networks using cell phones equipped with cameras. A cell

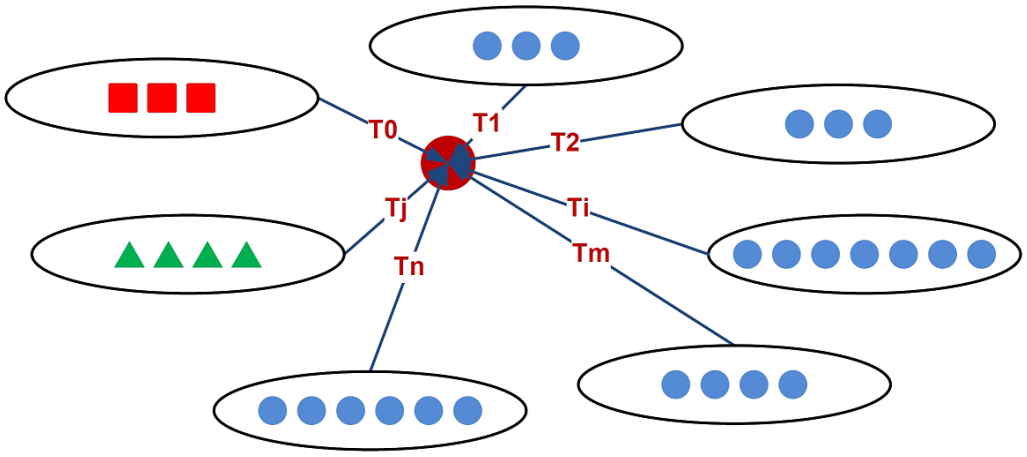

Figure 3.1: Multimodal Data Fusion. Measurements at successive time instances from different data sources with multiple modalities (i.e., radio, video, audio) are processed to localize and track a mobile user.

phone equipped with a camera in such networks acts as a real-time position sensor. From a single shot, taken by a cell phone, sent to a mobile switching center, a relative position is extracted. This rough estimate of global position is augmented with a estimated position from monocular vision. It results in geographical coordination (longitude and latitude) of the user with higher accuracy. In cellular networks, cell phones frequently report received signal strength values from 3 to 7 communication towers to mobile switching centers for both calling and non-calling users. This information can provide a rough estimate of mobile user location through multilateration. Another monocular vision based estimation of position is obtained upon receiving a camera shot from a user by template matching within available local visual databases (e.g., Google) to benefit from their geographical information tags (longitude and latitude).

Particle filtering uses a probabilistic transition model between a state and a probabilistic observation model to determine the next state using observed evidences. It represents the posterior distribution by a set of weighted samples (particles) that estimate probability distribution. Most of the existing localization algorithms did not consider the actual situation of multiple measurements and cannot be directly applied. The novelty in this approach is in utilizing particle filters for both localization and fusion of bimodal data.

State-space models originated from dynamic control theory. A state-space model usually consists of two sets of equations, the system (dynamic) equation: $x_{t+1} = f_t(x_t, u_t)$ and

the observation (output) equation: $y_t = h_t(x_t, v_t)$, where the mappings $f_t : R^m \times R^p \rightarrow R^m$ and $h_t : R^m \times R^q \rightarrow R^k$ are assumed to be known. The system equation models the dynamics of state variables and the observation equation models the observed state variables. For a linear Gaussian state-space model, the well-known Kalman filtering approach provides optimal estimates for state variables based on the information from the two sources, the dynamic equations and the observations. However, for nonlinear and non-Gaussian state-space models, it is quite challenging to estimate (as well as filter, smooth and forecast) the state variables and model parameters. In general, the state x_k is frequently evaluated from the conditional probability density function $p(x_k|y_{1:k})$ and a set of measurements $y_{1:k} \equiv \{y_1, ..., y_k\}$ up to sample k by the Chapman-Kolmogorov equation:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \tag{3.1}$$

For every measurement y_k, the posterior state probability density function can be updated via the Bayes rule:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_y|y_{1:k-1})}{p(y_k|y_{1:k-1})} \tag{3.2}$$

where $p(y_k|y_{1:k-1})$ is a normalizing term. This equation can be written in recursive mode as:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|x_{1:k-1})}{p(y_k|y_{1:k-1})} p(x_{1:k-1}|y_{1:k-1}) \tag{3.3}$$

The analytical solution to the above equations is intractable. So, many computational methods such as Sequential Monte Carlo (SMC) approaches (A.K.A Particle Filters) and Gibbs sampler techniques have been developed for nonlinear and non-Gaussian state-space models in the past decades [34, 24].

3.1 Multimodal State Recovery

The proposed data fusion and localization method in this chapter uses signal-strength measurements to augment with position information extracted from monocular vision. For

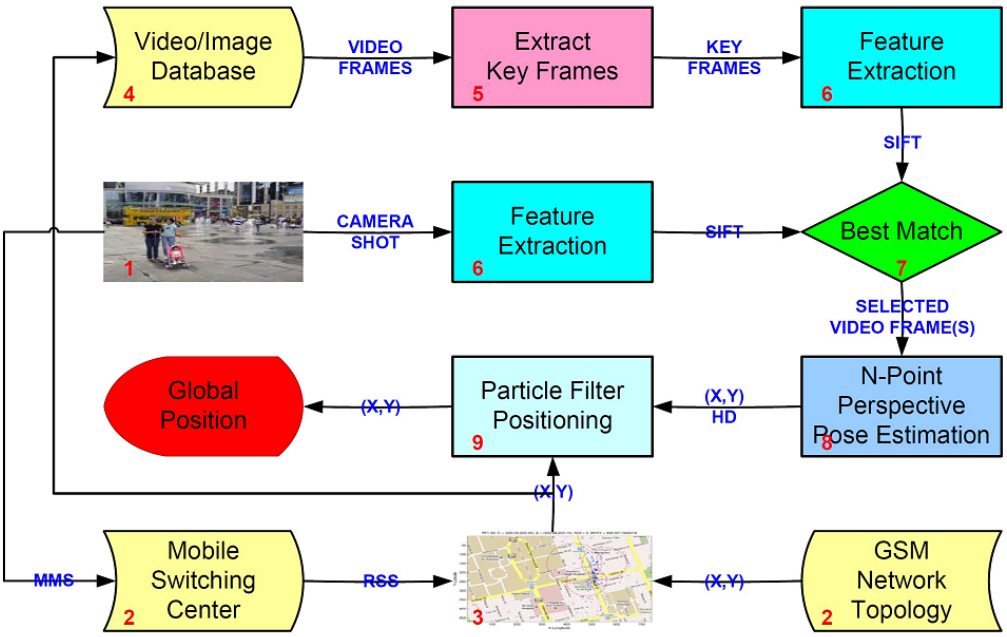

Figure 3.2: Bimodal localization in cellular networks. Extracted user position from signal strength measurements is augmented by estimated distance between a mobile user and landmark from monocular vision.

estimation and fusion of bimodal measurements, the particle filter is employed to find the position of mobile users in areas with inconsistent navigation signals. Most of the conventional approaches use the mathematical models to build a position vector at the first step. A tracking method like the Kalman filter can process this position vector later. Figure 3.2 is an overall view of the system. In this proposed method, we need to have access to local visual databases and maps with geographical tags (i.e., Google maps) and Network Measurement Reports. Some channel models require other information on the antenna type, antenna direction and beam width.

3.1.1 Monocular Pose Estimation

The proposed method starts from a single camera shot taken by a mobile user in the network. This shot is then used to search within the available visual databases to find a match with a video frame or image. The first step in video processing is key frame

Figure 3.3: Key frames selection in reference video to avoid ill-conditioned problem and have maximum camera motion between frames.

Figure 3.4: Feature extraction for frame matching. Both images have been divided to M sub-frames with minimum N features. Hamming Distance is used to find the best match between frames of local video and taken users' shot.

selection. If there is not enough camera motion between two frames, the computation of the epipolar geometry will be an ill-conditioned problem. To avoid this problem, we select frames with the highest camera motion between frames while still maintaining the ability to match with previously localized geography.

The first frame in a video shot is always chosen as a key frame (I_1). The next key frame I_2 must have at least M common interest points with the previous one. As a result, the key frame I_{n+1} has at least M common interest points with I_n and at least N common interest points with I_{n-1}. This ensures that there are enough matching points between key frames to calculate camera motion (Figure 3.3).

After key frame selection the next step is template matching. It is a fundamental part

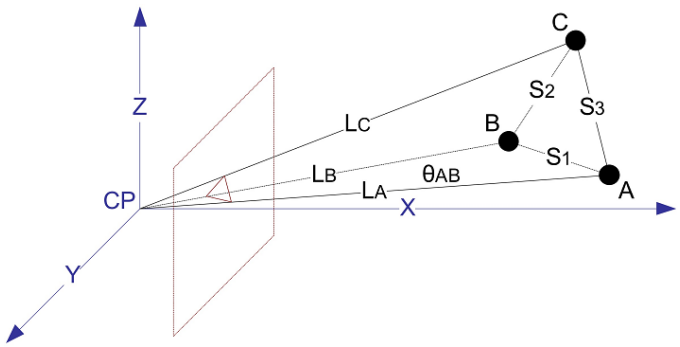

Figure 3.5: Perspective-3-Point pose estimation. The problem is to determine the lengths L_A, L_B, and L_C from 3D point positions A, B, and C [48].

of many problems of object or scene recognition. The extracted features for matching must be invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint and well localized in both the spatial and frequency domains. A scale Invariant Feature Transform (SIFT) method is used to extract features for template matching between the shot take by the cell phone user and reference video/image sequences [49].

Figure 3.4 shows one iteration of feature extraction and the matching process. In this shot, frames have been divided to 16 sub-frames before feature extraction [85].

Given the perspective projection of three points constituting the vertices of a known triangle in 3D space, it is possible to determine the position of each of the vertices. This is a well-known problem in computer vision (Figure 3.5) with six different solutions [48]. All of them benefit from the law of cosines (3.4).

$$S_1^2 = L_A^2 + L_B^2 - L_A \times L_B \times \cos\theta_{AB} \tag{3.4}$$

Solving Perspective-n-Point (PnP) problems correspond to find the roots of (3.5). By solving this equation, unknown distances L_A, L_B and L_C can be defined from known values s_1, s_2 and s_3 and perspective angles θ_{AB}, θ_{BC} and θ_{AC}.

$$G_0 + G_1 x + G_2 x^2 + G_3 x^3 + G_4 x^4 = 0 \tag{3.5}$$

The accuracy of monocular pose-estimation depends on several factors such as changes in the environment like occlusion by parked cars, moving people, and trees that can arise from shot to shot [48].

3.1.2 Inconsistent Signal Strength Measurements

Signal attenuation is the most accessible and affordable measurement to use in estimating the node-to-node distance; however, it suffers from noisy and inaccurate measurements or delays due to fading channels. Most of the current range-based localization methods first estimate the distance based on the empirical channel models and then infer location. Trilateration is the most basic and intuitive way for positioning [121]. This method computes a node's position via the intersection of three circles. In real-world applications, the distance estimation inaccuracies as well as the erroneous position information of reference nodes results in an infinite set of possible positions.

Furthermore, when a larger number of reference points are available, we can use multilateration to determine the node position and an over determined system of equations must be solved. Usually, over determined systems do not have a unique solution ($Ax = b$), but can be easily solved using traditional methods (e.g., the least squares method). The presented method in (3.6) has been employed to derive the radius of the circles in which a mobile user moves with the highest probability [97, 60].

$$(x - \frac{k^2 x_2 - x_1}{k^2 - 1})^2 + (y - \frac{k^2 y_2 - y_1}{k^2 - 1})^2 = (\frac{kD}{k^2 - 1})^2$$

$$\log_{10}(k) = \frac{A_1 - A_2}{10n} + N(0, \frac{2\sigma^2(1-\rho)}{100n^2}) \tag{3.6}$$

In this equation, signal attenuation (A_i), the path loss exponent (n), the distance ratio ($k = \frac{d_1}{d_2}$), the distance between two base stations (D), the standard deviation ($\sigma^2 = 2.2 - 8.3$) and a correlation coefficient of shadow components ($\rho = 0.3 - 0.8$) are the input parameters [71, 87].

For GSM localization, we need to extract the required field from the Network Measurement Report (NMR) database. Table 3.1 lists the typical fields in the NMR file. Primary positioning in many localization algorithms requires this information.

To extract the real distance between Mobile Stations (MS) and Base Transceiver Systems (BTS), base station physical geometry is required too. Some of the communication models require more information about the antenna type, antenna direction, beam width and so on. This information is available in another database (Table 3.1).

Figure 3.6 is a typical distribution of base stations in a cellular network.

Network Management Report	
Timestamp	Time stamp of event
Cell	The CGI of the current serving cell
Event	Type of report
IMSI	IMSI
IMEI	IMEI
Key	Current encryption key associated with the MS
TA	Current Timing Advance in microseconds
trxTxLevel	TRX power reduction in dB
Downlink	Downlink receive level in dBm
Bsc	PredictionModel
Cell	AntennaUserLabel
Ncc	BeamDirection
Bcc	HeightAGL
BCCH	MechanicalAntennaTilt
GsmCellLocation	Long
GsmCellIdentity	Lat
GsmCellFrequencyBand	HorizontalBeamwidth
ERP	AntennaType

Table 3.1: Network Management Report.

3.1.3 Bimodal data fusion using particle filters

The dynamics of the system can be described by the system equation:

$$x_k = f(x_{k-1}, w_k) \tag{3.7}$$

where $x_k \in R^{n_x}$ is the system base state (position), and $w_k \in R^{n_w}$ is the state noise, with $k \in N$, N is the set of natural numbers. The measurement equation has the form of:

$$y_k = h(x_k, v_k) \tag{3.8}$$

where $y_k \in R^{n_z}$ is the observation, and $v_k \in R^{n_v}$ is the measurement noise. Functions $f(.)$ and $h(.)$ are nonlinear in general. It is assumed that the observations are taken at discrete time points from either radio or visual signals. For the bimodal localization process,

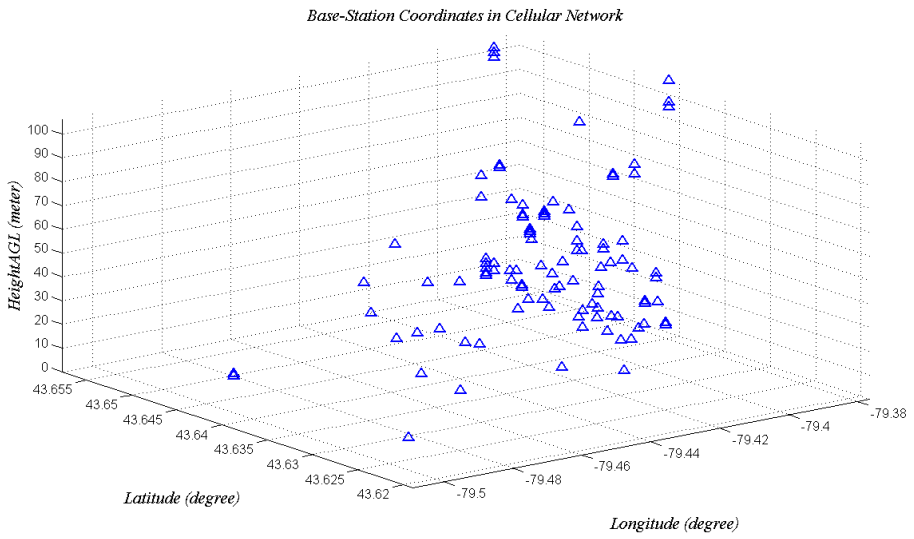

Figure 3.6: A typical distributions of communication towers in a cellular network.

we have utilized hybrid particles to fully characterize the state x_k which is frequently evaluated from the conditional probability density function and a set of bimodal measurements $y_{1:k} = \{y_1, y_2, ..., y_k\}$ up to time instant k.

As here we have two different sources of data that change over time, the sampling method must be adaptive, e.g., when we receive a camera shot, we can extract the position from visual data with higher accuracy than RSS measurements at the same time. Sometimes we have occlusion in the cell phone camera photo or scene itself, or the local video is not up to date. In this case, we must adjust the particles' weights to minimize inaccurate measurements effects. By this adjustment, the weight of a particle will be proportional to the normalized combination of two likelihoods (3.9) when measurements are available from both sources:

$$p(y_k|x_k^i) = \alpha_1 \times p(y_k^1|x_k^i) + \alpha_2 \times p(y_k^2|x_k^i) \tag{3.9}$$

In this equation, α_1 and α_2 values (modality confidence coefficients) can be set based on a probabilistic or deterministic modal model. In the simplest form, they can be set based on the quality of measurements. In some environments, the quality of measurements drop significantly due to propagation channel (e.g., obstruction, reflection, diffraction). We may assign lower weights at these cases. Figure 3.8 shows the effect of confidence coefficients

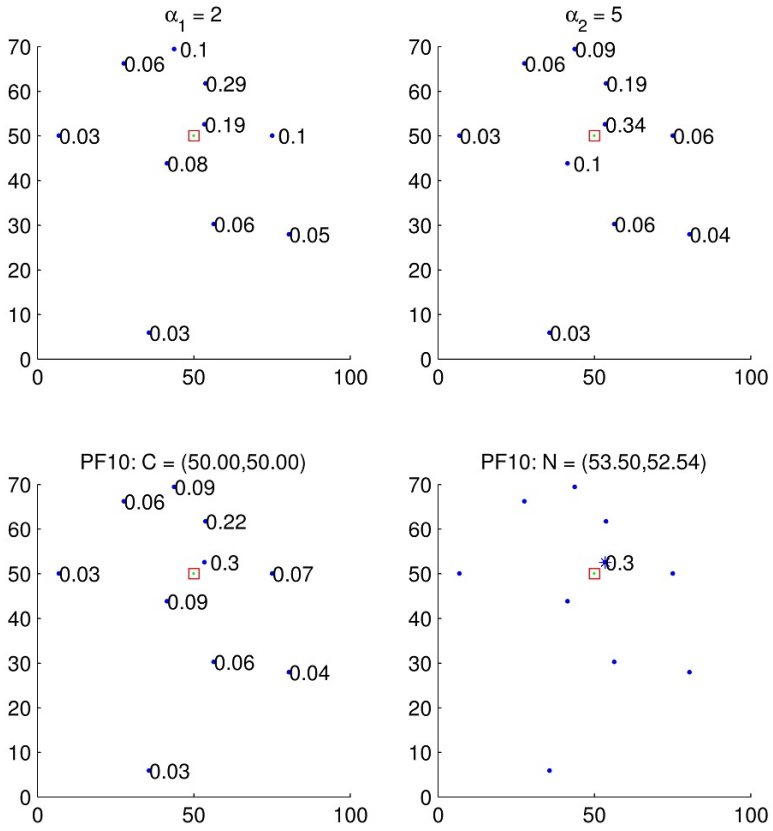

Figure 3.7: Particle fusion by adaptive weighting of samples from radio ($\alpha_1 = 2$) and visual ($\alpha_2 = 5$) sources. The confidence coefficients can be defined based on the environmental condition.

α_1 (for radio particles) and α_2 (for visual particles) on estimation error for a typical case of uncertainty ratio between measurements. In case of data lacking for a modal, these values may tend to zero. It can result in long converging time. Figure 3.7 shows the hybrid particles' weights for a single iteration of our filtering scheme with 1000 particles at maximum.

For data fusion and filtering, a general particle filtering method has been applied in four stages for time instances $k = 0, 1, 2, \ldots$ as follows:

- for $i = 0, 1, 2, \ldots, N$, sample
  $x_k^i \sim p(x_k | x_{0:k-1}^i, y_{0:k})$ that $x_{0:k}^i \equiv (x_{0:k-1}^i, x_k^i)$

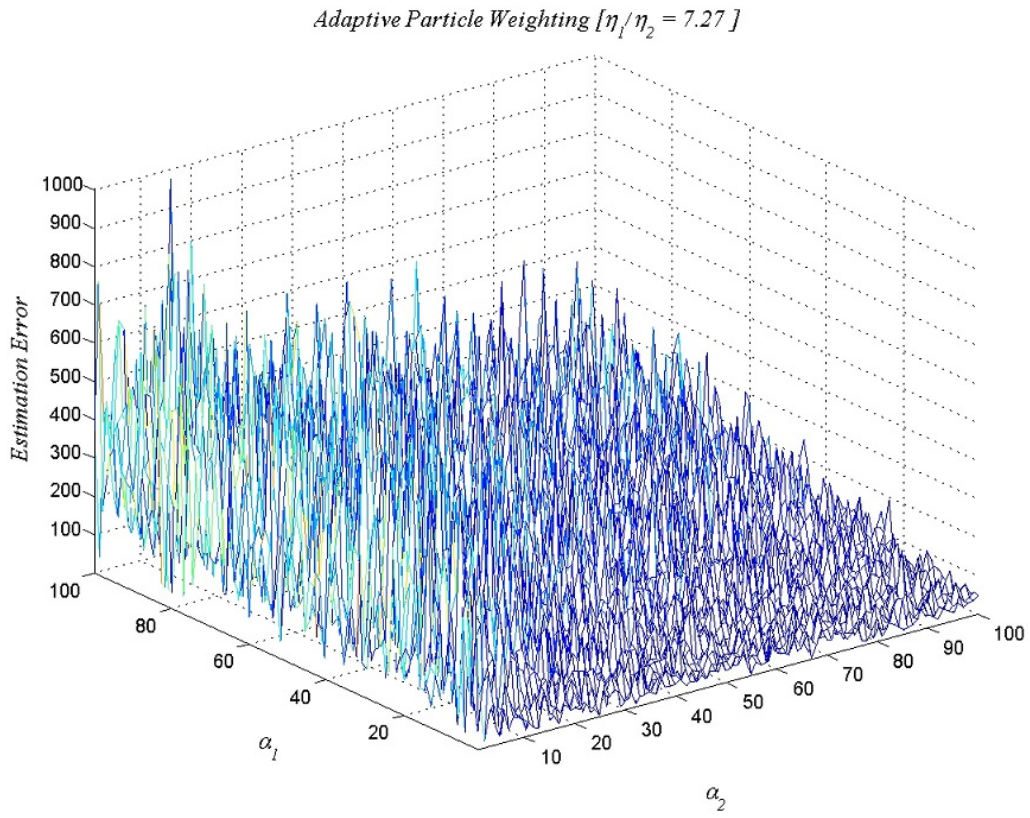

Figure 3.8: Estimation error vs. modal confidence coefficients α_1 (radio) and α_2 (visual) for measurement uncertainty ratio $\eta_1/\eta_2 = 7.27$. Higher confidence coefficient value for the 2^{nd} source causes lower estimation error.

- for $i = 0, 1, 2, \ldots, N$, compute weights for each source
  $\hat{w}_k^i(s) = \hat{w}_{k-1}^i(s) \times \frac{p(y_k^s|x_k^i)p(x_k^i|x_{k-1}^i)}{p(x_k|x_{0:k-1}^i, y_{0:k}^s)}$
- for $i = 0, 1, 2, \ldots, N$, compute fused weights
  $\hat{w}_k^i = \alpha_1 \times \hat{w}_k^i(1) + \alpha_2 \times \hat{w}_k^i(2)$
- for $i = 0, 1, 2, \ldots, N$, normalize weights
  $\tilde{w}_k^i = \hat{w}_k^i / \sum_{j=1}^N \hat{w}_k^j$

where $\hat{w}_k^i(s)$ is the non-normalized weight of sample i at time k for source s. As

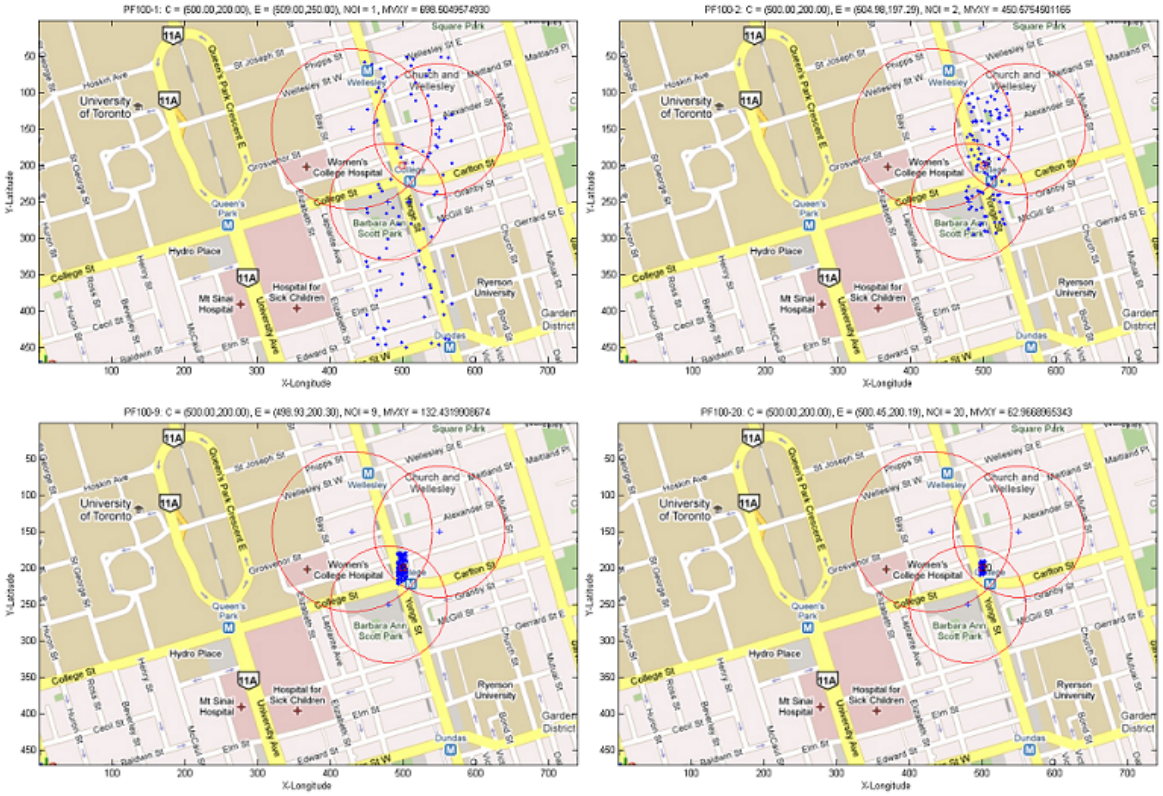

Figure 3.9: Particle filtering evolution with 50 hybrid particles after 1, 2, 9 and 20 iterations around the primary user position, located by trilateration in a cellular network.

a performance metric, the estimate $\hat{I_N}(f_n)$ of the posterior expectation $I_N(f_n)$ can be obtained using the equation:

$$\hat{I_N}(f_n) = \sum_{i=1}^{N} f_n(x_{0:n}^i) \frac{\hat{w}_k^i}{\sum_{j=1}^{N} \hat{w}_k^j} \tag{3.10}$$

In this working flow, two main stages can be distinguished: transition and resampling. During the transition, each particle is weighted according to the state model in the presence of noise. Then in the update stage the particle's weight is re-evaluated based on the recent measurement. A resampling procedure eliminates particles with small weights and replicates them with higher weights.

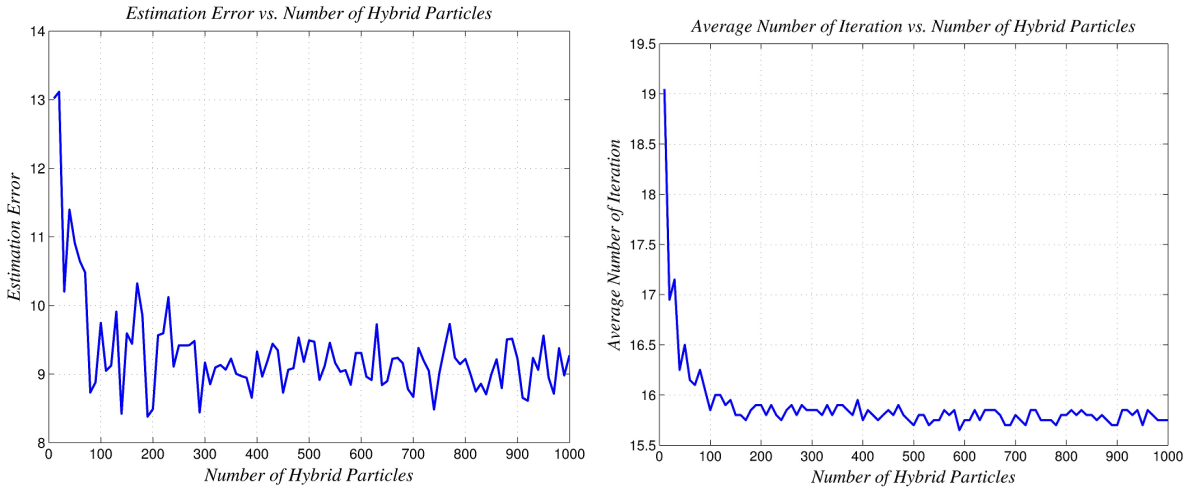

Figure 3.10: Position estimation error and number of iterations versus number of hybrid particles with the same error-mean threshold value. Number-of-particles plays a key role in system convergence time and the accuracy of localization.

Figure 3.9 shows the particles' distribution after a few iterations. It is a combination of extracted positions from available measurements on received signal strength and camera shots.

3.2 Experimental Results

We have developed a simulation framework to simultaneously implement, simulate and evaluate the proposed multimodal data fusion and localization methods under the same platform. This platform has several adjustable parameters that can affect the overall performance and computation cost. These parameters include camera calibration, video frame resolution, number of features in template matching process, number of partitions in video frames, window size in template matching, number of anchors in multilateration, modal confidence coefficients, number of particles and error-mean threshold. In the following results, the original image (shot and video frames) have been divided into 16 partitions. The minimum number of features in each partition is set to 12 (as Figure 3.4). By computing the Hamming Distance (HD) between feature points in partitions, we find the best match of the camera shot with a video frame. These HD values can be used directly for computing the posterior probabilities in particle filters. The system may fail for some reasons

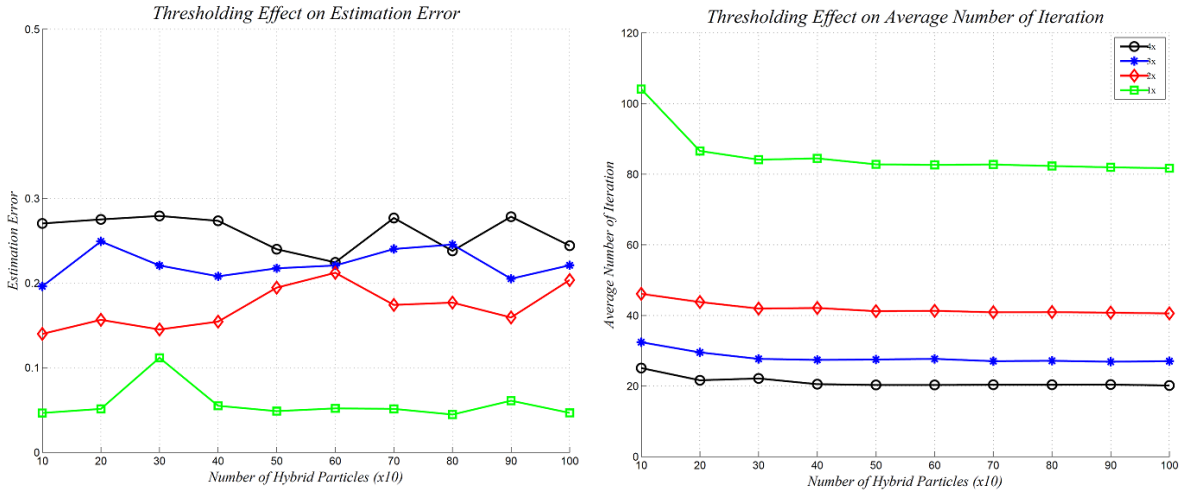

Figure 3.11: Estimation error and number of iterations vs. number of particle for different error mean thresholds: 100 (square), 200 (diamond), 300 (asterisk) and 400 (circle).

including high occlusion, especially in regions with high moving objects, i.e., people and cars. In such scenes, many of the feature points assigned to moving objects are misleading for the frame-matching algorithm.

Figure 3.10 depicts the estimation error versus the number of particles in the system as well as the average number of iterations versus the number of particles. The number of particles plays a key role in system convergence time along with the accuracy of localization.

The required number of particles depends on the problem itself and quality of acquired data. The system may converge with fewer particles if we provide high-quality data samples.

Estimation error can be investigated against the confidence coefficients α_1 (radio particles) and α_2 (visual particles) for a different measurement uncertainty ratio (i.e., $\eta_1/\eta_2 = 7.27$). A higher confidence coefficient value for the second source causes lower estimation errors. However, the confidence coefficient depends on the data quality itself and no reliable metrics exist to describe that. The frequency of camera shots has the least impact in generation of particles and even with a single camera shot we can start to localize the mobile user.

Figure 3.11 shows the effect of the error-mean threshold on the estimated error. Higher error thresholds (performance metric) result in low accuracy but fast convergence times.

3.3 Chapter Summary

In this chapter, we proposed a novel multimodal data fusion approach that applied for localization of mobile users in cellular networks. First, a series of vision-based algorithms applied to derive a preliminary mobile user position from monocular vision to augment with estimated location in a cellular network. These algorithms have been used for feature extraction, feature matching, perspective-n-point depth inversing. In the next step, particle filters employed to fuse bimodal radio-visual position estimations via signal strength and monocular vision measurements.

This approach improves the accuracy of signal strength-based localization methods by augmenting it with visual information. We show that fusing frequent inaccurate measurements with low-frequent accurate ones may increase the performance of localization significantly. This fusion can take place at different processing stages of mobility modeling.

The main limitation of the proposed method is the fact it requires an up to date video capture of the environment. Although, the experiments show that the proposed method is robust against environmental changes, it may not be suitable for highly changing environments, for example, in environments with parked cars, moving people, trees, and buildings that their visual characteristics (i.e., shape, color, brightness) vary across video instances. Occlusion can have a significant impact on the system performance.

Chapter 4

State estimation from Sparse and Inconsistent Measurements

Mobility models in CPS demand high degree of accuracy in state (i.e., position, velocity, acceleration) estimation of mobile agents, especially, when sparse measurements are available due to some restrictions in data acquisition. Accurate state estimation of mobile agents has long been considered as a challenging task in wireless networks. In this chapter, we investigate the sensor state estimation problem where a target sensor measures inconsistent signals as received-signal-strength or time-of-arrival from anchor sensors with known locations, whereas target sensor mobility state must be estimated. We know that even in large-scale wireless sensor networks, information is relatively sparse compared with the number of sensors. In such networks, the state estimation problem can be recast as a sparse signal recovery problem in the discrete spatial domain from a small number of linear measurements by solving an under-determined linear system. Exploitation of compressive sensing theory enables us to recover sparse signals from far fewer samples than the Nyquist sampling rate. The proposed method uses a few inconsistent measurements to estimate the mobile agent's state over a non-symmetric spatial grid. In this method, an ℓ^1-norm minimization program is used to recover the state of mobile agent in WiFi networks. The performance of the proposed method is evaluated through simulations with both synthetic and real measurements. Figure 4.1 is a sketch of problem that we tackle in this chapter. There are several constraints on choosing a state estimation method (e.g., computational power and environmental conditions). A simple solution is by using the Global-Positioning-System (GPS) that provides highly accurate location estimation. Such accuracy is affected by a number of factors, including satellite positions, noise in the radio signal, atmospheric conditions and natural barriers in the signal path. Unfortunately, in

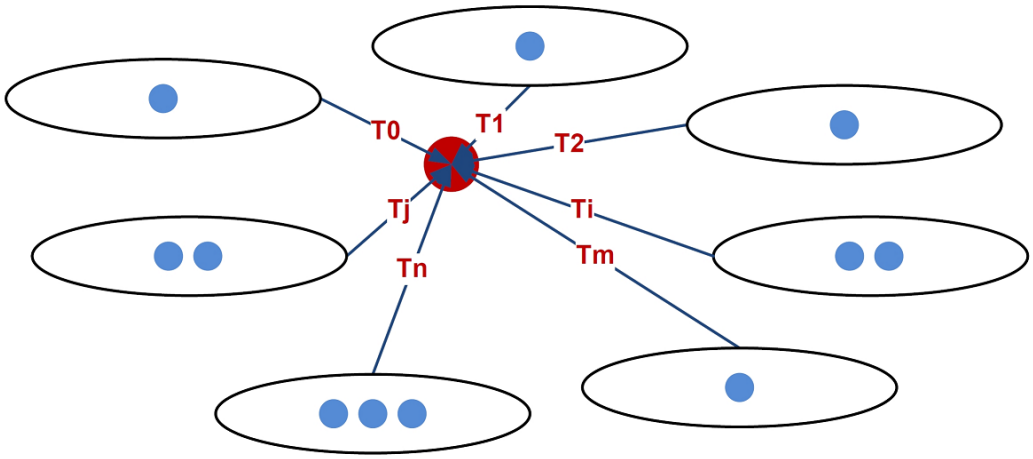

Figure 4.1: Processing sparse and inconsistent Measurements. Sparce measurements at successive time instances in the same modality (i.e., TOA, RSS or POA) are processed to recover the state (i.e., position) of a mobile agent.

dense urban areas or indoors, buildings can mask the received signal and in these cases the localization accuracy drops considerably. To overcome these problems, other localization techniques like anchor-based methods have been proposed where a few nodes (called anchors) have known locations while the remaining nodes need to estimate their positions [118]. Most of the proposed methods for wireless device localization use the anchor-based localization schemes. There is extensive research on static or dynamic state estimation that have recently been dominated by Bayesian techniques (i.e., Kalman filters, Particle filters, random finite sets) with multi-modal (i.e., radio, video, audio) data [24, 34, 106]. The proposed state estimation method in this thesis can be applied when the number of measurements are not enough (sparse) to estimate the user's state. In such cases, we can process a set of measurements to estimate the states of multiple users simultaneously. This method can be integrated with other localization approaches to augment them with compressive sensing unique features.

Compressive-Sensing, also known as compressed sensing, compressive sampling and sparse sampling, is a procedure for reconstructing a signal utilizing the prior sparse or compressible knowledge. The field has existed for several decades, but recently has been intensively considered again [32, 16, 99, 18] and as a consequence, it has been admired to improve the performance and robustness of state estimation methods from sparse information [21, 40]. In CS-based methods, the mobile device state can be estimated by finding the sparsest solution of an under-determined linear system. It has been proven

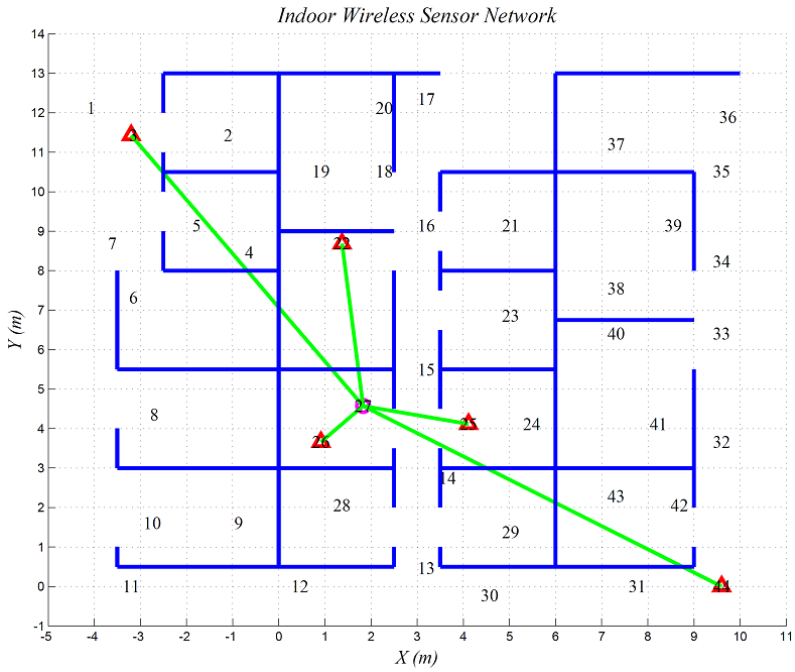

Figure 4.2: Mobility state estimation. Available measurements from five sensors are processed to estimate the target location. Sensors have distributed in $14m \times 13m$ office area. Measurement area is partitioned by $1.8m$-height cubicle walls with hard-partitioned offices, external glass windows, and cement walls on the outside of area.

that ℓ^1-norm minimization results is an optimal solution for this problem [33]. A complete list of proposed methods and solutions is available on [57].

Our proposed mobility estimation method is a compressive sensing approach to apply on sparse and inconsistent signal measurements where an accurate state estimation is not applicable due to sparse number of measurements. The estimation problem is formulated as a sparse approximation of a sparsifying matrix (dictionary) in which its elements are measurements of received signal (e.g., received-signal-strength or time-of-arrival) at discretized points or grids as Figure 4.2.

4.1 State Estimation through Compressive Sensing

The proposed state estimation method in this chapter utilizes CS theory to estimate wireless device position in cyber physical systems via noisy RSS or TOA measurements. Consider an isotropic medium with N reference points (grid), where a wireless device (target) is located in it with M available measurements from other devices (sensors) or reference points (anchors), while the number of measurements is much smaller than the grid size ($M \ll N$).

Recent theoretical results in CS have developed a series of algorithms to recover a signal from highly incomplete information via a small set of incoherent projections. The results also state that a sparse vector $x \in R^N$ can be recovered from a small number of linear measurements by solving an under-determined linear system:

$$y = Sx + \epsilon, \quad s.t. \;\; x \in R^K \;\; and \;\; K \ll N \tag{4.1}$$

The solution of the above problem is called the sparse solution. It is clear that the above problem can be solved in a finite time. However, there could be C_N^K combinations of non-singular linear systems to be solved. An undetermined system of linear equations has zero or infinite solutions that form an affine space [33]. The sparsest solution of (4.1) can be achieved by searching the minimum ℓ^p-norm in N-dimensional space, defined as:

$$\| x \|_p := \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p} \tag{4.2}$$

In specific cases, the ℓ^1 and ℓ^0 norms of a vector are defined as the summation of absolute values of vector elements and number of non-zero elements respectively:

$$\| x \|_1 := \sum_{i=1}^{N} |x_i| \quad , \quad \| x \|_0 := \sum_{i=1}^{N} \mu(x_i) \tag{4.3}$$

where

$$\mu(x) = \begin{cases} 0 & \text{if } x \text{ is zero} \\ 1 & \text{if } x \text{ is non-zero} \end{cases}$$

Under suitable conditions, minimizing the ℓ^1-norm is equivalent to minimizing ℓ^0-norm, but the former is computationally more tractable than the latter, where the latter one is very sensitive to measurement noise. Although there are many algorithms presented in literature to find the minimums in both norm spaces, the signal processing community usually uses the ℓ^1-norm to describe a signal with just a few components [31, 66, 82].

Most of the proposed solutions to find the minimum utilize linear programming (LP) or second-order cone programming (SOCP) approaches to solve the problem. One of the most successful solutions presented in the literatureknown as Basis Pursuits (BP)seeks the minimum ℓ^1-norm distance (Manhattan Distance) by linear programming:

$$\min_x \| x \|_1 \quad s.t. \quad y = Sx \tag{4.4}$$

In our work, we have used ℓ^1-magic toolbox which has implemented signal recovery procedures in MATLAB [57]. In this package, the linear and quadratic problems are solved using a generic path-following, primal-dual method and a generic log-barrier algorithm respectively [13]. Among the proposed solutions in this package, we have selected ℓ^1-optimization for signal reconstruction. With the above assumptions, the state estimation problem can be recast as a sparse signal recovery problem in the discrete spatial domain. According to the CS theory, the K-sparse representations x of sensor location can be recovered from noisy measurements y that satisfy this equation:

$$y = \Phi\Psi x + \epsilon \tag{4.5}$$

where $\Psi_{N\times N}$ is the sparsifying basis (dictionary), $\Phi_{M\times N}$ is the measurement matrix and ϵ is the measurement noise. The $\Psi_{N\times N}$ dictionary is made by signal readings on the grid points. Thus, Ψ_{ij} is the signal measurement reading on grid point i received from the grid point j.

For simulation purposes based on the RSS information, the sparsifying frame $\Psi_{N\times N}$ can be generated using the radio propagation channel model:

$$\Psi_{ij} = p_{ij} = p_0 - 10n_p\log_{10}(d_{ij}/d_0), \quad 1 \leq i, j \leq N \tag{4.6}$$

where p_{ij} is the received signal power in decibel milliwatts (dBm), d_{ij} is the real transmitter-receiver distance, d_0 is the reference distance for the antenna at far field with

received power p_0 and is the path loss exponent n_p. For TOA based simulations a similar equation can be used where t_{ij} is the propagation delay between nodes with speed of light c.

$$\Psi_{ij} = t_{ij} = d_{ij}/c, \quad 1 \leq i, j \leq N \tag{4.7}$$

To apply CS theory and solve the problem by linear programming the $\Psi\Phi$ must hold the Restricted Isometry Property [21, 16]. Since the sparsifying basis Ψ and the measurement matrix Φ are usually coherent, an orthogonalization step must be applied first [40]. Considering the sensing matrix $S = \Phi\Psi$, (4.5) can be recast to have an orthogonal basis T as:

$$TS^{\dagger}y = TS^{\dagger}Sx + TS^{\dagger}\epsilon \tag{4.8}$$

where T is the orthogonal basis for S (i.e., $TT' = I$) with pseudo-inverse $S^{\dagger}$. Thus, (4.5) is replaced by (4.9) creating a simple ℓ^1-minimization problem.

$$z = TS^{\dagger}y = Tx + \epsilon' \tag{4.9}$$

Figure 4.3 is a snapshot from the proposed localization process using CS for target number 4 in our simulation platform. This sensor has four RSS measurements y from sensors number 3, 10, 35 and 44 in a low-SNR regime. The estimated signal has 3 major components above a predefined 0.9 threshold. The estimated target location x has distance error 163cm. Figure 4.4 is the similar snapshot of the same target for TOA measurements.

The performance of the method depends on several parameters including the measurement noise ϵ, so in a recovered signal, we may find non-zero components C_i around the exact index of a target that must be considered to increase the performance of the system.

We have used the normalized values of theses coefficients to locate the target precisely. A threshold (thr) has been defined on the summation of normalized components C_i to select n largest components. Then an iterative method is applied to select m out of those n components ($2 \leq m \leq 3$) with minimum cumulative distance error overall. Eventually, the values of the selected components are processed to estimate the location of the target.

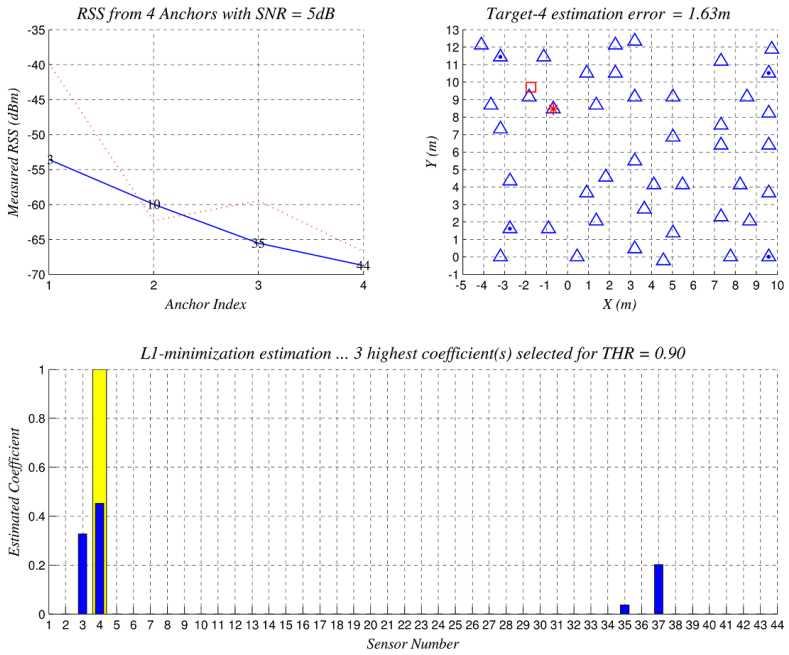

Figure 4.3: State estimation of sensor 4 (target) using RSS measurements from sensors $3, 10, 35$ and 44 (anchors) in low-SNR ($snr = 5dB$) and high-threshold ($thr = 0.9$) scenario. Three largest coefficients of recovered signal x are used to estimate target location ($err = 163cm$).

$$\underset{i=1}{\overset{\arg\min \|C_j\| \;\; s.t. \;\; 1 \leq j \leq n}{\sum}} C_i^s x = b \;\; s.t. \;\; \sum_{k=1}^{n} C_k > thr \tag{4.10}$$

An additional refinement scheme must be applied to increase the system performance, if we do not map the target extracted location to one of grid points.

4.2 Experimental Results

In this section we provide details on the experimental evaluation of the proposed method using real RSS or TOA measurements in an office environment. Our simulation setup was developed in MATLAB using the measurement data provided by [91]. The test environment

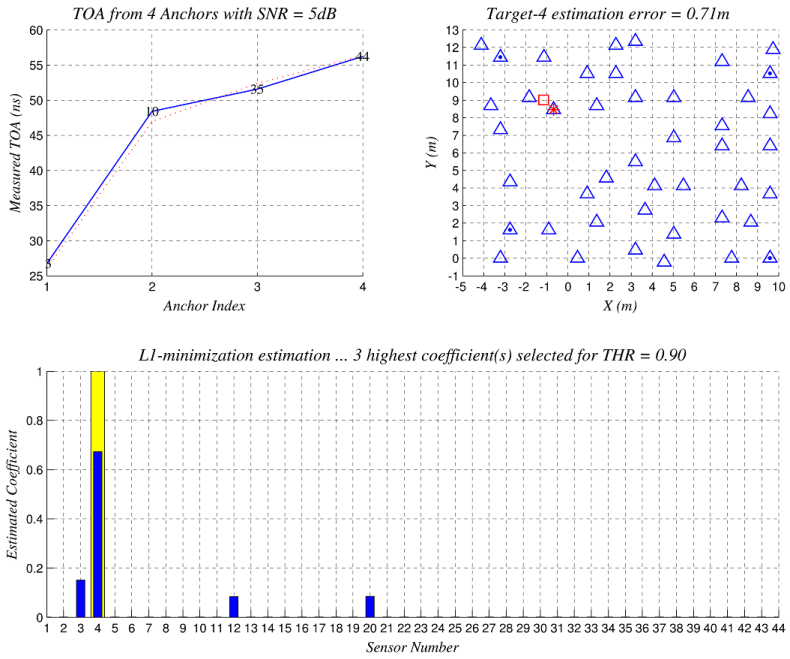

Figure 4.4: State estimation of sensor 4 (target) using TOA measurements from sensors $3, 10, 35$ and 44 (anchors) in low-SNR ($snr = 5dB$) and high-threshold ($thr = 0.9$) scenario. Three largest coefficients of recovered signal x are used to estimate target location ($err = 71cm$).

is a partitioned office with forty-four sensors located in identified coordinates within a 14m×13m area. Two sets of measurements are available for RSS and TOA signal types. The transmit power was 10mW at center frequency 2.443GHz and Signal-to-Noise ratio was maintained greater than 25dB to compensate for noise and ISM-band interference. For TOA measurements, both TX and RX are synchronized by 1 pulse per second (1PPS) signals. The time base standard deviation (≥2ns) was achieved in these measurements. All devices are in range of each other, so a total of 44×43×5=9460 measurements are taken.

Figure 4.2 shows this test-bed for simulations. Targets and anchors are selected in a random way from 44 possible reference points. The sparsity matrix $\Psi_{44\times44}$ (dictionary) has already been provided through the field measurements.

Figure 4.5 plots the distribution of measurements in both RSS and TOA dictionaries. Every entry in these matrices is an average of five measurements under the same conditions. Estimated bias $(t_{ij} - d_{ij}/c)$ for TOA measurements is around $10.9ns$, where t_{ij} and d_{ij} are estimated time and real distance between nodes i and j in test-bed [91].

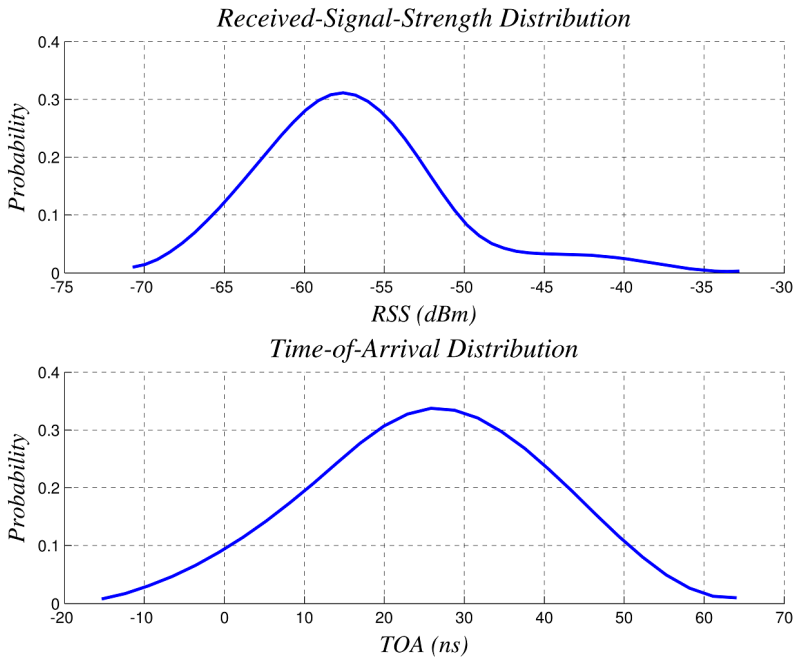

Figure 4.5: Probability distribution of 9460 ($44 \times 43 \times 5$) RSS and TOA measurements. Estimated bias delay for TOA measurements is around $10.9ns$.

If the estimated sensor location is mapped to one of the grid points in discrete spatial (no refinement process), (6.1) can be used as a metric to find the normalized distance estimation error of the target in performance evaluation, otherwise Euclidean distance $\| x_i - \hat{x_i} \|_2$ is a proper metric.

$$err = \frac{\| x_i - \hat{x_i} \|_2}{\max(\| x_i - x_j \|_2)}, \ \forall \ j \in [1 \cdot \cdot N] \tag{4.11}$$

The performance of the proposed scheme is compared to the multilateration localization method. Figure 4.6 plots the distance estimation error for all sensors in our test-bed from randomly selected anchors. The *snr* and *thr* are fixed and the same for both methods in this result.

Figure 4.7 shows the performance of the proposed method in the presence of noise for different signal-to-noise ratios. In these measurements the transmitted signal power is fixed at 10mW. As we expected, in high-SNR regimes, sensor location can be recovered perfectly

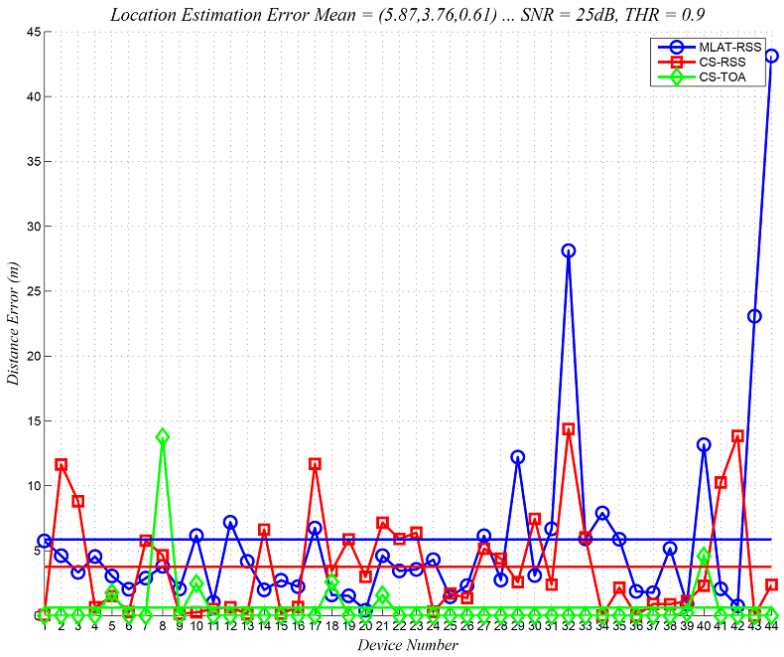

Figure 4.6: The performance of position state recovery. Multilateration vs. Compressive Sensing. The *snr* and *thr* are set to 25dB and 0.9. Average estimation error (*err*) is 4.96, 3.64 and 1.22 for the corresponding methods.

from a few measurements. In this scenario the main component in the recovered signal x is usually above the threshold (*thr*) itself.

Figure 4.7 also shows the performance vs. the different threshold values. In low-SNR regimes, low thresholds involve more non-zero components in the refinement process that increase error probability. An adaptive method can be developed to set the threshold proportional to the signal-to-noise ratio.

4.3 Chapter Summary

To provide a model of user mobility pattern in CPS, we shall solve the state estimation problem from sparse and inconsistent measurements in an efficient way. The motivation for the presented method in this chapter has been to verify the accuracy with which one can estimate the sensor state, utilizing compressive sensing theory when sparse measurements exist. This chapter began with problem formulation of state estimation from sparse and

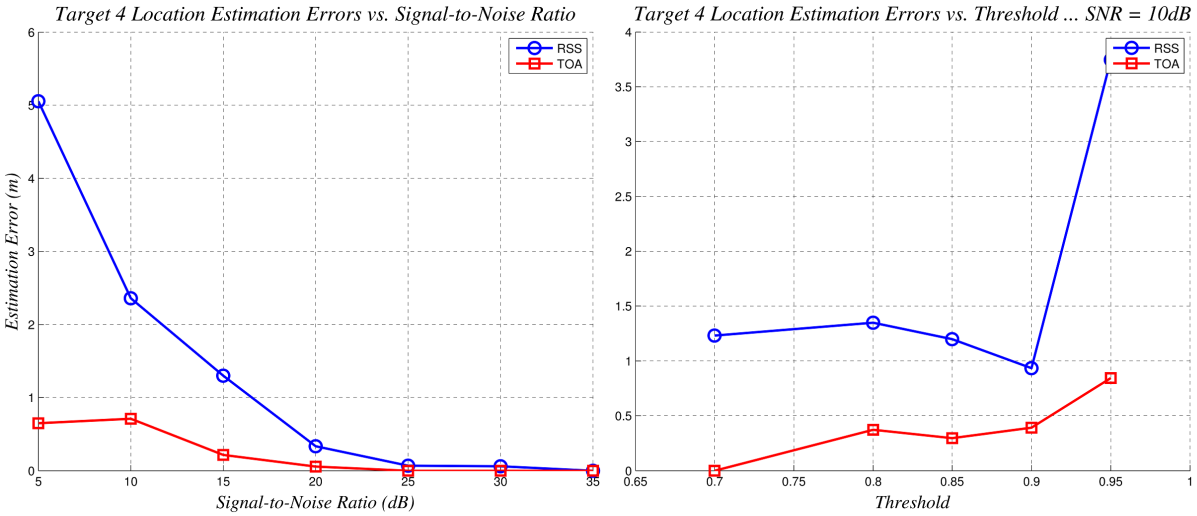

Figure 4.7: Estimation error in CS localization for different *snr* and *thr* values. The threshold and signal-to-noise values are set to 0.9 and 10dB respectively. The higher threshold values involve more coefficients in refinement process that make system unstable.

inconsistent measurements inside the compressive sensing framework. We found that the target state can be recovered precisely from sparse RSS and TOA measurements in the high-SNR regimes. In low-SNR regimes, the estimation from the TOA measurements shows higher performance than the RSS measurements that still suffer from a high fluctuation in signal amplitude. The performance depends on several parameters such as the signal-to-noise ratio, number of targets, number of measurements, linear programming procedure, sensor deployment and network topology. As we expected, noise power has the main effect on localization performance using RSS information. Constructing an adaptive sensing matrix is the most difficult part of the system that affects the performance directly. This method was proposed primarily to mobility state estimation in a CPS when a sparse number of measurements are available due to system restrictions. Applying this method to massive sensory data in wireless sensor networks may improve the resource utilization in many demanding applications.

Chapter 5

Kernel-Based Optimization for Mobility Modeling

Traffic estimation and prediction services play a key role in most cyber physical systems as intelligent transportation systems. The performance of such systems is heavily dependent on the availability of traffic flow information and the system's ability to analyze the different types of floating car data to derive reliable traffic flow estimates. In this chapter, we propose a parametric statistical approach to estimate the traffic flow from floating car data. The localized cumulative distance among Gaussian kernels is minimized through quadratic optimization to weigh the Gaussian mixtures located on the points-of-interest in traffic estimation processes. Road geospatial coordinates are employed as optimization constraints and as priori knowledge on kernel means and bandwidths. A probabilistic framework is used to extract system parameters by modeling the measurements by means of a generalized Gaussian density function. Computational complexity is relaxed by linear approximation of kernel weights and the effect of this process on performance is investigated. The proposed approach is applied to real measurements of floating car data obtained from cellphone mobility data. This approach can be augmented by other traffic sensory data, such as cameras and loop sensors. Figure 5.1 is a sketch of problem that we tackle in this chapter.

5.1 Mathematical Formulation

We proposes a kernel-based method to model the traffic flow at macroscopic level. A statistical approach is used to extract kernel parameters (i.e., mean and bandwidth) from

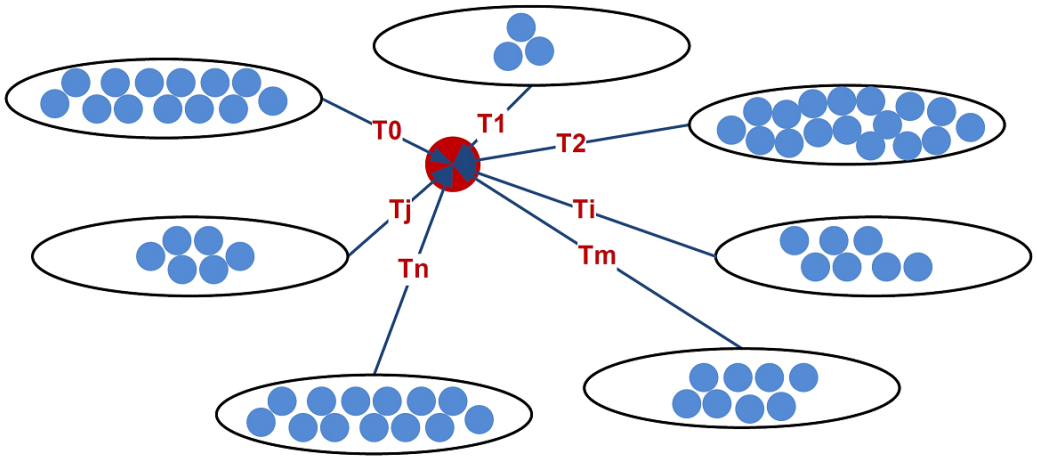

Figure 5.1: Processing massive sensory data. Collected position samples in the same modality at successive time instances are processed to model the density of traffic flow in CPS.

floating car data. In the rest of this section, we review the mathematical formulation of kernel-based densities in general form and the role of Localized Cumulative Distribution (LCD) function in comparing and finding the distance (similarity) between distributions.

In the parametric minimum distance estimation, the most important step is the quantification of degree-of-closeness between the sample data and the parametric model. This distance between the data and the model is usually described as a parametric function. In statistical inference, two broad types of distances between the distribution functions and the probability density functions are presented. The Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling distances are examples of first-type distances. The second-type of distances (i.e., density-based distances) includes the Hellinger, Kullback-Leibler divergence, the family of chi-square distances as a whole, and some other families like Bregman divergence and Burbea-Rao divergence [11]. While the parametric inference is the primary use of minimum distance estimation methods, it can be used as a test of goodness-of-fit. There is also fundamental difference between the two classes of applications. In parametric inference, robustness is a concern but the aim is to decrease the effect of inconsistent measurements with the model. In goodness-of-fit problem, we are interested in magnifying the small deviations from the hypothesized model to achieve higher performance.

The Kolmogorov-Smirnov general distance measurement ρ between distribution functions F and G is defined as:

$$\rho_{KS}(F,G) = \sup_{-\infty<z<+\infty} |F(z) - G(z)| \sqrt{\psi(F(z))} \tag{5.1}$$

where $\psi(u) = 1$ gives the Kolmogorov-Smirnov distance measure.

The weighted Cramer-von Mises distance is defined as:

$$\rho_{CM}(F,G) = \int_{-\infty}^{+\infty} (F(z) - G(z))^2 \psi(F(z)) dF(z) \tag{5.2}$$

where $\psi(u) = 1$ gives the Cramer-von Mises and $\psi(u) = [u(1-u)]^{-1}$ gives the Anderson-Darling distance measure.

The Kolmogorov-Smirnov test has long been used as a useful tool in various applications, e.g., approximation of distribution functions, goodness-of-fit test and test for normality. The Cramer-von Mises test is widely used in goodness-of-fit testing and parametric estimation. The Anderson-Darling distance test outperforms the other distance tests in this class for the normality test. In our approach we have used the Cramer-von Mises distance test for approximating the empirical data samples from massive sensory data (i.e., Floating Car Data).

Given a set of N_s independent and identically distributed (i.i.d.) random sample vectors $\vec{x}_i := [x_i^1, x_i^2, \ldots, x_i^N] \in R^N$ in N-dimensional domain, empirical probability density in the form of a mixture of Dirac distributions can be represented as the function f_D with normalized weights ω_i:

$$f_D(\vec{x}) = \sum_{i=1}^{|D|} \omega_i \delta(\vec{x} - \vec{x}_i) \tag{5.3}$$

This density function shall be estimated in the form of axis-aligned Gaussian mixtures in a N dimensional domain:

$$f_{GM}(\vec{x}) = \sum_{i=1}^{|GM|} \alpha_i \prod_{k=1}^{N} \mathcal{N}(x_k - \mu_{k_i}, \sigma_{k_i}) \tag{5.4}$$

$$\mathcal{N}(x - \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{5.5}$$

The above mixture contains N_p Gaussian components, one aligned to each of the N-dimensions with mean μ_{k_i} and standard deviation s_{k_i}. In our estimation, the Gaussian components are centered at data samples ($\mu_{k_i} = x_s^i$) while the standard-variation (bandwidth) of each dimension is fixed and determined as a priori ($\sigma_{k_i} = s$). By making these assumptions, the only unknown parameter in (5.4) is the kernel weights $\vec{\alpha} = [\alpha_1, \ldots, \alpha_M]$. These kernel weights shall be optimized later to minimize the distance between the empirical and estimated distributions.

In order to compare probability distributions or determine their similarity, several methods (e.g., Kolmogorov-Smirnov, Cramer-von Mises) have been proposed, but generally they are not suitable for multi-dimensional probability distributions.

5.2 Density Modeling in Transportation Systems

Intelligent Transportation Systems (ITS) are complex CPS that combine cyber aspects (i.e., wireless communication) with physical aspects (i.e., movement) in time and space. Current ITS implementations employ advanced technologies and modern management techniques to deliver important services, including Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS). The efficiency of such systems is highly demanded. Recent developments in Intelligent Transportation Systems have placed more importance on the management of existing resources to enhance performance and provide new services such as Dynamic Traffic Assignment (DTA) and Traffic Estimation and Prediction (TEP). In both services, traffic flow modeling is a key requirement.

Traffic congestion is one of the main issues to be addressed due to its impact on the economy and the environment. This persistent issue has been a global concern for the last few decades and effective solutions have yet to be witnessed. Recognizing that there are more vehicles being introduced daily into the already limited-capacity road network makes one realize that this issue will continue to worsen unless effective solutions are developed. The reduction of traffic congestion can be achieved through effective deployment of an infrastructure of fixed or mobile sensors, e.g., radars, cameras, loop detectors, navigation devices and even pedestrian crossing buttons to collect input data for traffic management. These devices usually measure a vast variety of traffic-related quantities (e.g., position, speed and vehicle counts) which can be employed in behavioral modeling and prediction of the traffic systems. Most of the fixed sensors are placed on highways and roads to report traffic statistics frequently. On the other hand, we also expect valuable streams of sensory data through distributed mobile sensors in moving cars, i.e., floating car data (FCD). The traffic congestion problem can be tackled by different mathematical (i.e., statistical

or probabilistic) methods. An extensive literature on the subject of statistical and probabilistic modeling exists with the most important probabilistic models being the Bayesian Networks, the Dynamic Bayesian Networks and the Hidden Markov Models because they can be also utilized for density estimation of sensory data samples [101].

The main objective of this work is to provide a statistical framework to model and process massive amounts of floating car sensory data to tackle the road traffic congestion problem. The proposed method is a kernel-based optimization method that minimize the distance between localized cumulative distributions of data samples [63]. Such probabilistic frameworks can be employed as a data fusion engine in traffic management with multiple modalities of sensory data. Kernel density estimation is a parametric way to estimate a probability density function and works as a data smoothing solution when inferences are made based on a finite number of data samples that have been used studied over the past decades in transportation systems [76].

In our approach, in order to compare the f_D and f_{GM} distributions, the Cramer-von Mises distance between their cumulative distributions is used. The localized cumulative distribution is an alternative representation of the cumulative distribution function obtained from the integration of probability density with symmetric kernels. Unlike the cumulative distribution, the localized cumulative distribution is unique, symmetric and well-defined for multi-dimensional probability distributions [47].

Given a multivariate N-dimensional random vector $\vec{x} \in R^N$ and the corresponding probability density function $f(\vec{x}) : R^N \rightarrow R_+$, the localized cumulative distribution $F(\vec{x}, b) : R_+^N \times R^N \rightarrow [0, 1]$ with the bandwidth $b \in R_+$ is defined as:

$$F(\vec{x}, b) = P(|\vec{x} - x| \leq \frac{1}{2} b) \tag{5.6}$$

The cumulative distribution function can be generalized based on a suitable kernel $K_b(\vec{x}, \vec{m}) : R^N \times R_+^N \rightarrow [0, 1]$ centered at $\vec{m}$ with bandwidth b as:

$$F(\vec{m}, b) = \int_{R^N} f(\vec{x}).K_b(\vec{x}, \vec{m}) d\vec{x} \tag{5.7}$$

If we consider only separable Gaussian kernels with mean $\vec{m} = [m_1, m_2, \ldots, m_N]$ and identical bandwidth b, we can rephrase the kernel function $K_b(\vec{x}, \vec{m})$ as:

$$K_b(\vec{x}, \vec{m}) = \prod_{k=1}^{N} \mathcal{N}(x_k - m_k, b) \tag{5.8}$$

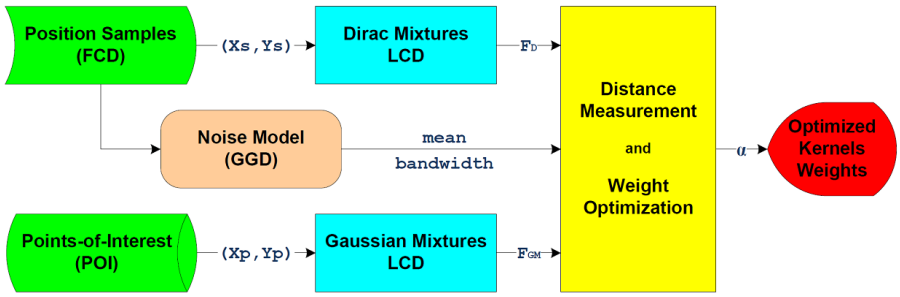

Figure 5.2: Proposed method: probability distribution functions of collected position samples from Floating Car Data (FCD) and points-of-interests in traffic modeling are approximated with Gaussian mixtures. The Cramer-von Mises distance between Localized Cumulative Distributions (LCD) of these probability densities is minimized to extract the optimized weights of Gaussian mixtures. From optimized weights, road densities including the points-of-interests can be interpolated.

After computation of the localized cumulative distributions, the distance between probability density functions (5.3) and (5.4) shall be minimized to find the optimized weights of the Gaussian mixtures (α_i) in (5.4). Eventually, from this optimized distributions, we extract the densities for the points-of-interest in our local map.

5.3 Non-Parametric Kernel Weighting

We propose a statistical approach that can be employed in processing massive position sensory data, collected from floating cars to model road traffic density in transportation systems. A series of i.i.d. position samples have been processed to extract optimized weights of Gaussian mixtures with predefined kernel parameters (i.e., mean and variance). The optimized kernels build the density distribution function of the samples that can be easily converted later to standard traffic metrics.

Figure 5.2 is the block diagram of the proposed method. Probability density functions of collected position samples from Floating Car Data (FCD) and Points-of-Interest (POI) in transportation networks are approximated by Gaussian kernels with predefined mean and variance parameters, then the distance between their localized cumulative distributions is minimized to extract the optimized weights of the Gaussian mixtures.

As our field data usually comes from 2D measurements (e.g., longitude and latitude), for the rest of this chapter, only separable Gaussian kernels in two-dimensional domain

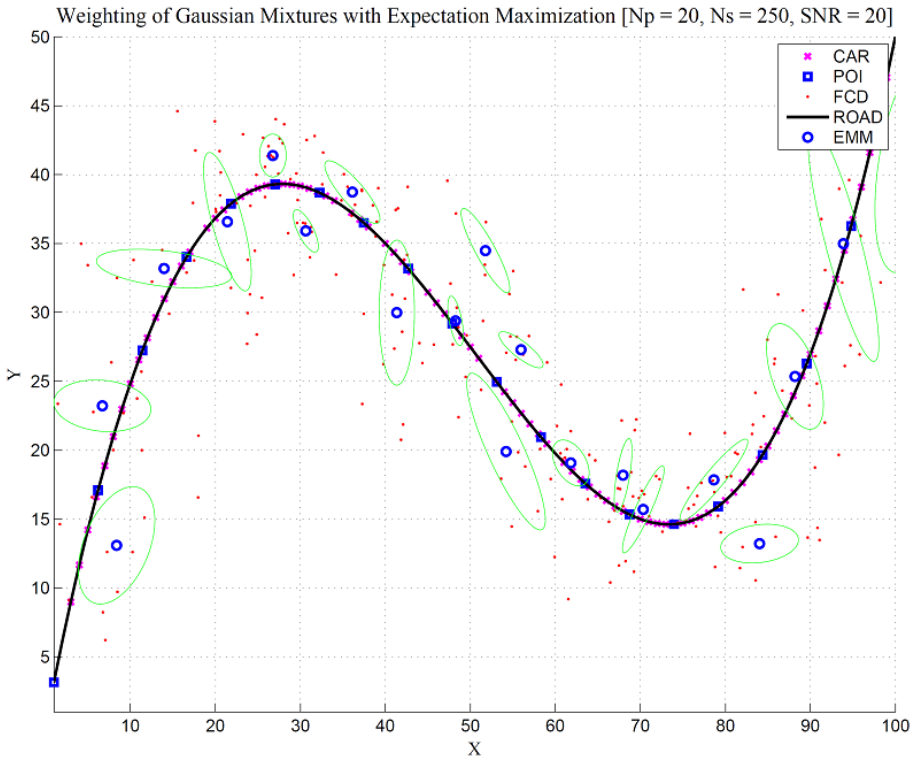

Figure 5.3: Snapshot of sample distributions for road density estimation through Expectation Maximization for Gaussian mixtures. The collected position samples from driving cars are approximated by a mixture of Gaussian kernels to extract the road density in transportation network. The kernel means and variances are extracted for the same points-of-interests.

$(N = 2)$ with mean $\vec{m} = \{m_x, m_y\}$ and bandwidth $\vec{s} = \{s_x, s_y\}$ are considered. With this assumption, (5.4) can be rephrased as:

$$f_D(x, y) = \frac{1}{N_s} \sum_{i=1}^{N_s} \delta(x - x_s^i)\delta(y - y_s^i) \tag{5.9}$$

$$f_{GM}(x, y) = \sum_{i=1}^{N_p} \alpha_i \mathcal{N}(x - x_p^i, s_x)\mathcal{N}(y - y_p^i, s_y) \tag{5.10}$$

As a snapshot of empirical data, Figure 5.3 shows distributions of cars as well as points-

of-interest and derived position samples from cars (floating car data) in traffic modeling. A Gaussian kernel can be considered around each of the points under normal condition. In order to compare the probability distribution functions, the localized cumulative distribution is extracted first from an empirical data set. (5.11) and (5.12) are corresponding localized cumulative distributions of (5.9) and (5.10) respectively, in which, b and s are bandwidth and standard deviation of Gaussian kernels.

$$F_D(\vec{m}, b) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathcal{N}(x_s - m_x, b)\mathcal{N}(y_s - m_y, b) \tag{5.11}$$

$$F_{GM}(\vec{m}, b) = \sum_{i=1}^{N_p} \alpha_i \times \mathcal{N}(x_p - m_x, \sqrt{s_x^2 + b^2}) \times \mathcal{N}(y_p - m_y, \sqrt{s_y^2 + b^2}) \tag{5.12}$$

After finding the localized cumulative distributions of f_D and f_{GM}, the modified Cramer-von Mises distance between them is determined as:

$$D = \int_{R^+} \omega(b) \int_{R^N} (F_{GM}(\vec{m}, b) - F_D(\vec{m}, b))^2 d\vec{m}db \tag{5.13}$$

Distance D in (5.13) is only dependent on the kernel weights α_i in (5.10) that shall be minimized later [63]. It is obvious that the kernel weights are limited to one ($0 \leq \alpha_i \leq 1$) with a summation equal or less than a constant that can be adjusted based on the empirical sample distribution. For any optimization process, we usually have a regularization factor (R) in terms of optimization parameter to skip from the local minimums that can be defined as:

$$R = \int_{R^N} f(\vec{x})^2 dx \tag{5.14}$$

Considering distance measurement D and regularization term R, a new term $Q := D + c \times R$ can be defined quadratic in parameter α. The constant $c \in R$ is a trade-off coefficient that reflects the regularization belief in the optimization process. Simply, the optimization problem can be formulated as a quadratic programming in terms of α:

$$\hat{\vec{\alpha}} = \arg\min Q \ s.t. \ 0 \leq \alpha_i \leq 1 \tag{5.15}$$

Regularized distance Q is represented in closed forms [63] as:

$$Q := \vec{\alpha}^T P_1 \vec{\alpha} - 2\vec{\alpha}^T P_2 + c \times \vec{\alpha}^T P_3 \vec{\alpha} \tag{5.16}$$

Elements of matrix P_1 are calculated as:

$$P_1^{ij} = \frac{\pi b^4}{s^2 + b^2} \times \exp\{-\frac{1}{4}\frac{(x_p^i - x_p^j)^2 + (y_p^i - y_p^j)^2}{s^2 + b^2}\} \tag{5.17}$$

In addition to P_1, the matrix P_2 can be represented as:

$$P_2^{ij} = \frac{1}{N_s}\sum_{j=1}^{N_s} \frac{2\pi b^4}{s^2 + 2b^2} \times \exp\{-\frac{1}{2}\frac{(x_s^i - x_p^j)^2 + (y_s^i - y_p^j)^2}{s^2 + 2b^2}\} \tag{5.18}$$

and P_3 as:

$$P_3^{ij} = \frac{\pi b^4}{s^2} \times \exp\{-\frac{1}{4}\frac{(x_p^i - x_p^j)^2 + (y_p^i - y_p^j)^2}{s^2}\} \tag{5.19}$$

The additional constraints on parameter α_i may enforce the optimization process to result in a convex combination of weights. Finally, the Gaussian mixtures with optimized weights construct the density estimates around the points-of-interest. The density values for other points on the road can be extracted from primary estimates through this interpolation.

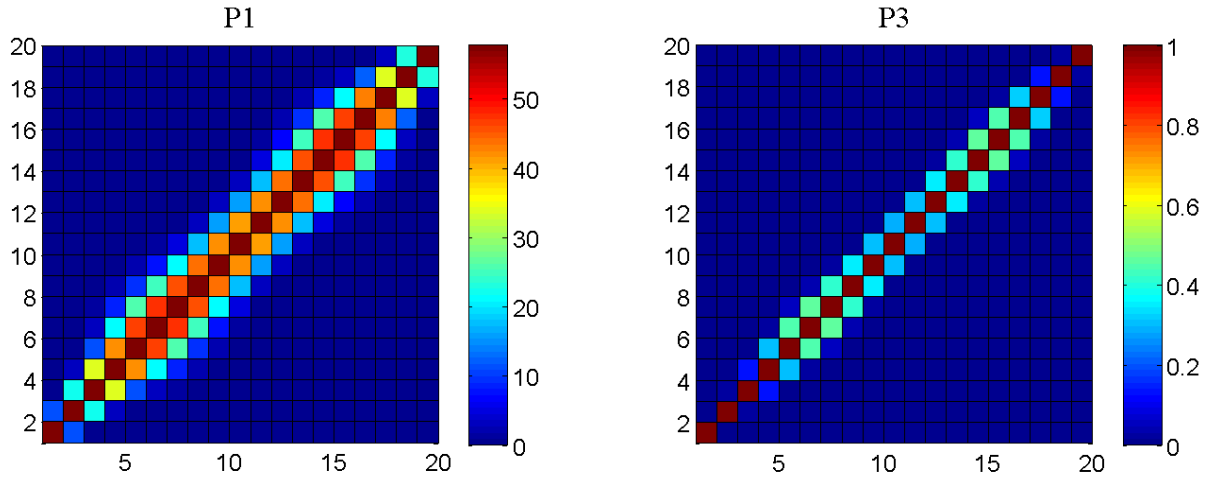

Figure 5.4: Sparse P1 and P3 matrices for highways. The off-diagonal elements of these matrices tend to zero where the points-of-interests are spatially distributed on a highway.

5.3.1 Linear Approximation

To reduce the computational costs of the method, we may investigate the degradation due to the linear approximation instead of the quadratic one. In case of largely distributed means $\vec{m}$, the off diagonal elements in P_1 and P_3 matrices tend to zero (Figure 5.4). In this case the distribution distances can be approximated based on the P_2 matrix that is a linear combination of the kernels' weights.

$$D_L = \sum_{i=1}^{N_p} \alpha_i \sum_{j=1}^{N_s} \exp\{-\frac{1}{2}\frac{(x_s^i - x_p^j)^2 + (y_s^i - y_p^j)^2}{s^2 + 2b^2}\} \tag{5.20}$$

By this assumption, we can rephrase the (5.16) as (5.20) and find optimized weights through linear optimization.

5.3.2 Expectation Maximization

Expectation Maximization (EM) is one of numeral methods to evaluate multi-variate functions [28]. In order to compare the performance of the proposed method with other methods in existence based on the Gaussian Mixtures, we have utilized the EM algorithmproposed

by Dempster, Laird and Rubin [28]to find the mean and variance of the mixtures. The EM algorithm has become a popular tool in statistical estimation problems involving incomplete data or similar problems such as mixture estimation.

The EM algorithm is a general method to find an explanation of the observed data $Y = Ym, m = 1, \ldots, M$ in terms of a parametric description of data X. Formally, we seek a maximum-likelihood estimate of the parameter vector θ:

$$\hat{\theta} = \arg\min_{\theta} \log p_Y(Y|\theta) \tag{5.21}$$

while p_Y is related to the complete-data X by:

$$p_Y(Y|\theta) = \int_{X(Y)} p(X|\theta)dX \tag{5.22}$$

where $X(Y)$ denotes all values of X for which $Y(X) = Y$. Since we do not have the full data X to compute the solution (5.22) directly, we maximize its expectation $E[\log p(X|\theta)]$. This expectation is taken over the probability distribution governing X, which is determined by the known values Y and the probability density function describing the unobserved portion of X.

Unfortunately, we do not have the parameter vector θ that defines the probability distribution governing X as this vector is exactly what we set out to find in the first place. Therefore, we use an estimate of it and iteratively improve it. Let us define a function Q that expresses the sought expectation of the likelihood as a function of the parameters θ that we are trying to estimate given the observed data Y and a current estimate $\hat{\theta}$ of the parameters:

$$Q(\theta|\hat{\theta}) = E[\log p(X|\theta)|Y, \hat{\theta}] \tag{5.23}$$

This Q function will allow us to compute the expected log-likelihood of the complete data X for any parameterization θ, while the expectations are computed using a fixed probability distribution defined by the observed data Y and given a parameterization $\hat{\theta}$. If we use the Normal probability distribution $p_k(y|\mu_k, \sigma_k) = \mathcal{N}(y - \mu_k, \sigma_k)$, we can estimate the Gaussian Mixtures mean and variance by expectation maximization [77, 78].

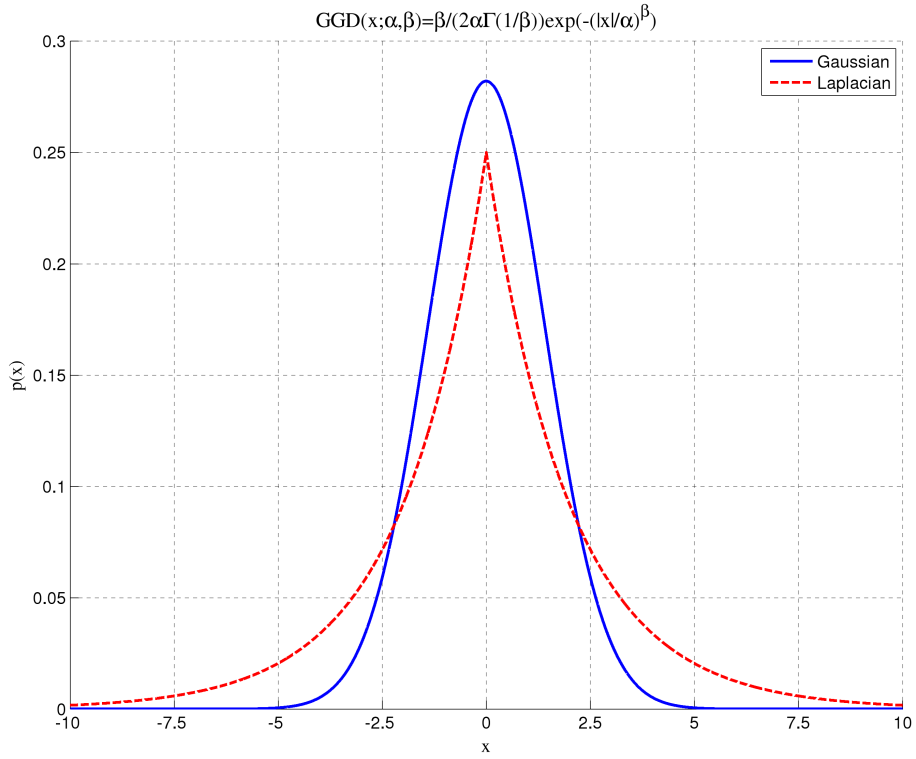

Figure 5.5: Generalized Gaussian Density (GGD) for Laplacian (red) and Gaussian (blue) special cases with the same standard-deviation ($\alpha = 2$).

5.3.3 Data Assimilation

As we find in (5.17) and (5.18), the measured distance D is very sensitive to the kernel's bandwidth (b). This parameter is itself dependent on geographical constraints that can be estimated from position samples (e.g., GPS logs). In order to have a proper estimate of kernel bandwidth parameter, we have utilized the Generalized Gaussian Density (GGD) function which is defined as:

$$p(x;\alpha,\beta) = \frac{\alpha}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^{\beta}} \tag{5.24}$$

where $\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt$ is the Gamma function and α models the peak-width (standard-deviation) of the Probability Density Function (PDF), while β is inversely proportional with the decreasing rate of the peak [30]. Sometimes α is referred to *scale*

parameter while β is called the *shape* parameter. In special cases when $\beta = 1$ and $\beta = 2$, it models the Laplacian and Gaussian PDFs respectively. Figure 5.5 shows the special cases of GGD with the same standard deviation. It is shown that, among classic statistical methods, the maximum-likelihood (ML) estimator gives the best results for GGD models in parameter estimation for both large and small samples. The likelihood function of the sample $\vec{x} = \{x_1, x_2, \ldots, x_N\}$ is defined as:

$$\ell(x; \alpha, \beta) = \log \prod_{i=1}^{N} p(x_i; \alpha, \beta) \tag{5.25}$$

where α and β are parameters to be estimated. It is shown that α and β parameters are the unique roots of the following maximum likelihood function:

$$\frac{\partial \ell(x; \alpha, \beta)}{\partial \alpha} = -\frac{N}{\alpha} + \sum_{i=1}^{N} \frac{\beta |x_i|^{\beta} \alpha^{-\beta}}{\alpha} = 0 \tag{5.26}$$

$$\frac{\partial \ell(x; \alpha, \beta)}{\partial \beta} = -\frac{N}{\beta} + \frac{N\Psi(1/\beta)}{\beta^2} - \sum_{i=1}^{N} \left(\frac{|x_i|}{\alpha}\right)^{\beta} \log\left(\frac{|x_i|}{\alpha}\right) = 0 \tag{5.27}$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function [116]. If we fix $\beta > 0$ then from (5.26) the parameter α can be estimated as:

$$\hat{\alpha} = \left(\frac{\beta}{N} \sum_{i=1}^{N} |x_i|^{\beta}\right)^{1/\beta} \tag{5.28}$$

And at the same time (5.27) parameter $\hat{\beta}$ is the solution for the following equation:

$$\Theta(\hat{\beta}) = 1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} + \frac{1}{\hat{\beta}} \log\left(\frac{\hat{\beta}}{N} \sum_{i=1}^{N} |x_i|^{\hat{\beta}}\right) - \left(\sum_{i=1}^{N} |x_i|^{\hat{\beta}} \log |x_i|\right) \left(\sum_{i=1}^{N} |x_i|^{\hat{\beta}}\right)^{-1} \tag{5.29}$$

The above equation can be easily solved by numerical methods (e.g., Newton-Raphson) effectively to find the GGD parameters for data modeling [30, 116].

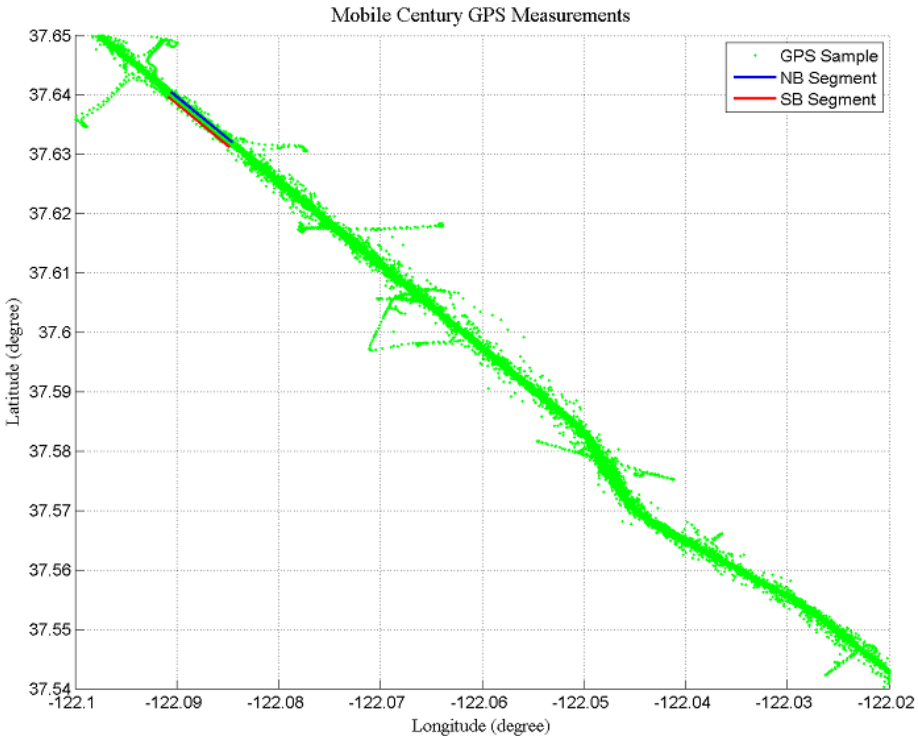

Figure 5.6: Stretch of highway I-880 CA, used in the Mobile Century experiment. The vehicles repeatedly drove loops of six to ten miles in length continuously for eight hours on four-lane freeway I-880 near Union City in the San Francisco Bay Area, California [52]. GPS samples collected in *Mobile Century* experiment. A segment of highway is selected (highlighted) to model its position samples by Generalized Gaussian Distribution.

5.4 Experimental Results

In order to evaluate the performance of the proposed method for road traffic modeling, a simulation framework was developed in MATLAB to process both the synthetic and realistic floating car data under different scenarios. To present robust statistics, all results are averaged over 100 runs for every experiment. The performance of the proposed method must be verified for each parameter individually (i.e., N_p, N_s, SNR, b, s, and c) for both synthetic and realistic data.

To illustrate the proposed method, we have employed it to model the collected floating car data from 200 GPS-enabled mobile phones in vehicles for an extended period of three months and the potential data [52]. Nicknamed the *Mobile Century* experiment, it involved

100 vehicles carrying GPS-enabled Nokia N95 phones. All drivers were instructed to drive normally with no other specific instructions. The vehicles repeatedly drove loops in lengths of six to ten miles continuously for eight hours on highway I-880 in the San Francisco Bay Area in California.

Figure 5.6 shows the collected GPS samples (FCD) in the *Mobile Century* experiment. To define the points-of-interest, highways can be divided by the concatenated segments between ramps (i.e., the exit or entrance ramp). We expect smooth density changes for each segment. Northbound and southbound traffic can be analyzed independently if we have enough accuracy (standard deviation of samples compared to road width) for position samples (i.e., FCD).

To extract the noise model of FCD in Figure 5.6 and investigating data accuracy, we may employ the equations (5.28) and (5.29) to extract the *scale* and *shape* parameters of GGD model.

Figure 5.7 shows distribution of collected GPS measurements as well as the extracted parameters for selected segments. The collected GPS measurements shown can be modeled with scale and shape parameters $\alpha = 4.65$ and $\beta = 1.17$ respectively. For simplicity, we have used the Gaussian distribution ($\beta = 2$) with the standard deviation $s = 1.96 \times \alpha \simeq 9$ to cover 95% of error probability. The noise model is independent from the physical characteristics of the roads but it is dependent on measurement noise and the propagation channel.

The bandwidth of kernel function (b) and noise model (s) have the highest influence on the performance of the system and distance D between the localized cumulative distributions. The ratio between these bandwidths (b/s) can be investigated as an impact factor on the density signal recovery error. The larger ratio value incorporates more position samples in the estimation of density for each point-of-interest. Figure 5.8 plots the effect of the ratio on normalized distance between cumulative distributions.

Besides the ratio between bandwidths, ratios can be independently changed to see their effect on the optimized weights of the Gaussian mixtures under similar SNR regimes.

An exhaustive test was also conducted to profile the SNR effect on optimized weights α_i and their variation under different SNR regimes. Notice that in the ideal case and with uniform distributions of FCD samples, for a system with ten points of interest, the optimized weights are equal to $\alpha_i = 1/N_p = 0.1$.

If we set the standard deviation s of the Gaussian Mixtures to extracted one from GGD modeling, we can plot the estimation error versus to the kernel bandwidth b. As shown in Figure 5.9 (left), it is monotonically increasing for quadratic optimization but decreasing

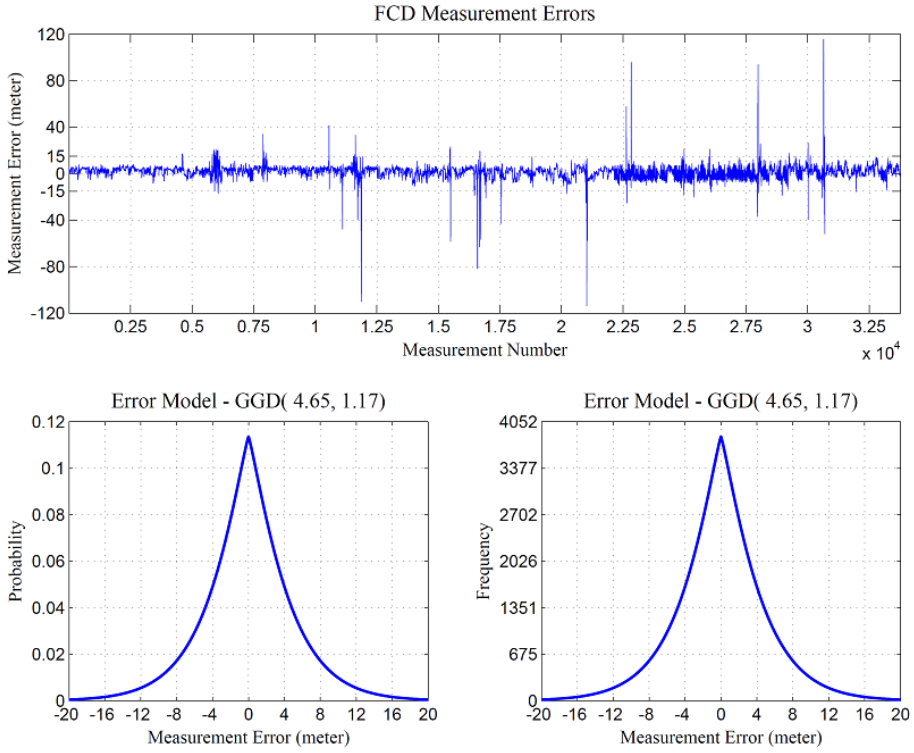

Figure 5.7: Estimated parameters of Generalized Gaussian Density $(\alpha, \beta) = (4.65, 1.17)$ from Floating Car GPS data samples in highlighted segment of Figure 5.6.

for linear approximation. There is also a trade-off factor c in (5.16) that increases the effect of the regularization term R in optimization. Increasing this factor shows our belief in regularization of the distance between LCDs. The magnitude order of regularization factor should be set in an adaptive way on kernel bandwidths themselves.

Figure 5.9 (right) also shows a snapshot of the weight optimization process (linear and quadratic) of the synthetic data in Figure 5.3 from 250 samples for 20 points-of-interests under normal and high SNR conditions. The variations are around the uniform values of weights ($N_p^{-1} = 0.05$). Linear approximation gives more smooth variation on weights α_i than a quadratic one.

The performance of the proposed method can be compared to similar approaches in literature. Some analytical and an experimental comparison between the utilized method and other parametric density estimation methods (e.g., EM, SVM) for Gaussian mixtures have been already provided in the literature [63, 64]. We compared our method with the

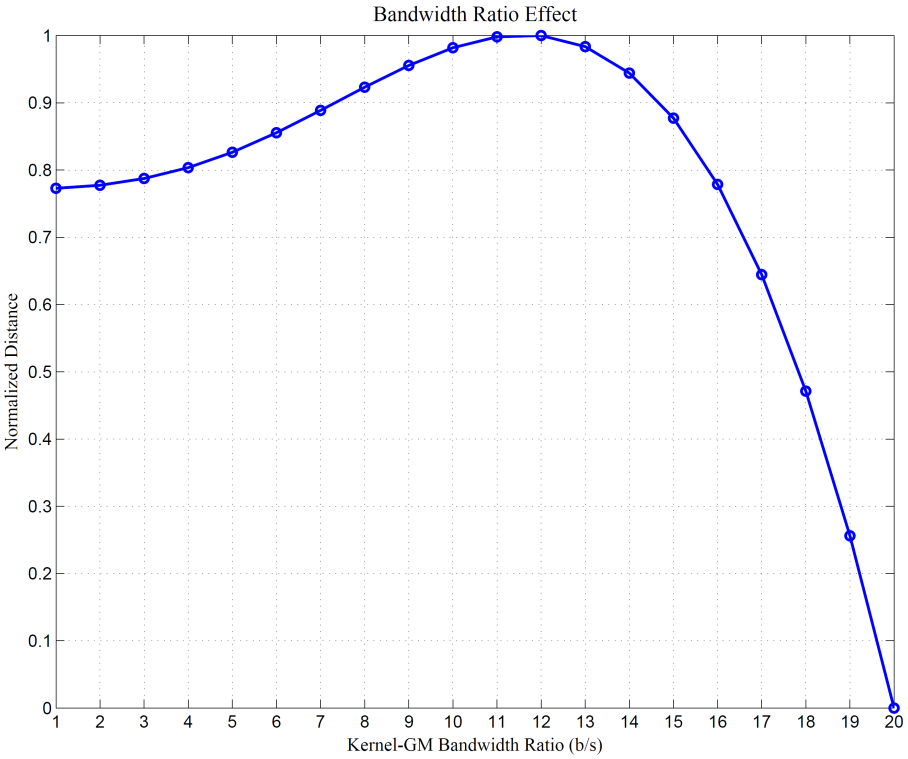

Figure 5.8: Normalized LCD distance between distributions (f_D and f_{GM}) after optimization. The ratio between kernel bandwidths has the highest impact on system's performance.

classic Expectation Maximization (EM) approach. Figure 5.3 shows the extracted Gaussian kernels (i.e., mean and variance) by EM from the position samples. The centers (means) of extracted Gaussian mixtures by EM are not necessary located on the road while we know that all samples originated from cars driving on the road.

Figure 5.3 also shows the optimized weights that are extracted by the proposed method (i.e., linear and quadratic) and the expectation maximization. The number of points-of-interest and number-of-kernels are set to 20 to have a fair comparison between kernel weights.

After finding the optimized weights for the Gaussian mixtures, we can estimate the road traffic density based on the Gaussian mixtures at every point-of-interest. Figure 5.10 shows the normalized density estimation for the same road.

The extracted kernels mean and variance are used to estimate the traffic density for points-of-interest for traffic estimation.

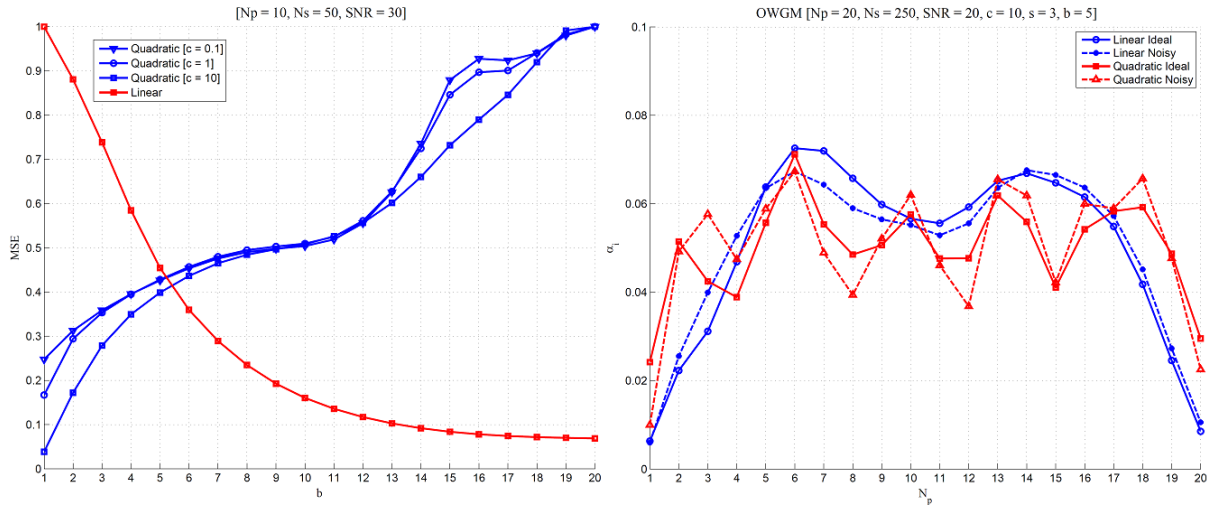

Figure 5.9: Kernel bandwidth effect. The plot shows the effect of bandwidth variation on signal recovery error in a high-SNR regime for linear and quadratic programs. The optimized weights through linear and quadratic minimization are plotted for 20 points-of-interests for both ideal and noisy off-road measurements.

The number of measurements has also one of the highest impacts on the performance of this system. Figure 5.11 plots the estimation error versus the number of available measurements in modeling and distance minimization. The plot reveals that even under different SNR regimes, eventually, our estimation error converges. The minimum number of required samples depends on system parameters as well as the spatial and temporal distributions of the FCD samples. We may need to process a higher number of available position samples in some traffic conditions, e.g., when the temporal or spatial distribution is sparse.

5.5 Chapter Summary

This chapter has examined the problem of traffic density modeling in transportation networks and proposed a kernel-based density optimization technique to resolve this problem. The weights of the Gaussian mixtures optimized through minimization of the Cramer-von Mises distance between the localized cumulative distributions of the derived position samples and the points-of-interest in traffic modeling.

The collected position samples of floating car data was modeled with Generalized Gaus-

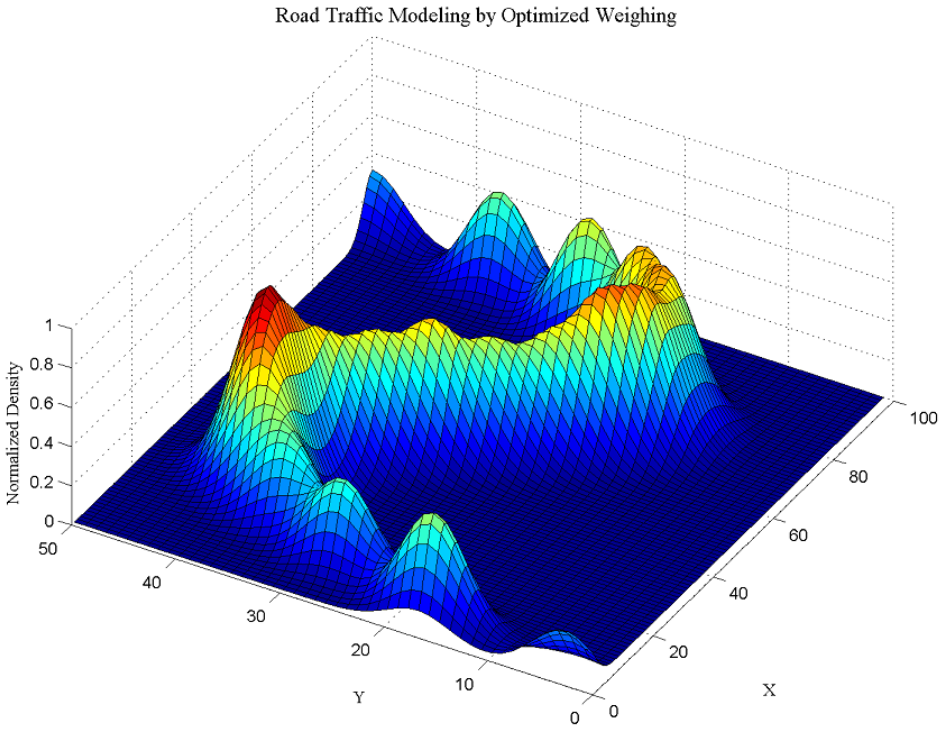

Figure 5.10: Road Traffic Modeling. Optimum kernel weights are extracted by minimizing the distance between localized cumulative distributions of position samples from floating car data and points-of-interest. The results are normalized for every road segment.

sian Density. Then, the distance between the localized cumulative distributions of the probability densities centered at floating cars positions and points-of-interests was computed. This distance was minimized through quadratic programming. From these optimized weights, the probability distribution function around new points-of-interests can be interpolated or estimated. The approach was applied to collected GPS data to evaluate its performance.

The performance of the system depends on many parameters including the bandwidth of kernels (b and s) as well as the trade-off factor (c) on regularization that can be defined in an adaptive way for future work. Having a spatial-temporal model also helps us in defining the kernels' bandwidth as a Priori in the optimization process.

The proposed method can be easily adopted to model the traffic evolution in urban areas under different traffic conditions. It can be also employed for congestion detection or movement pattern analysis indoors using the collected position samples from hand-held

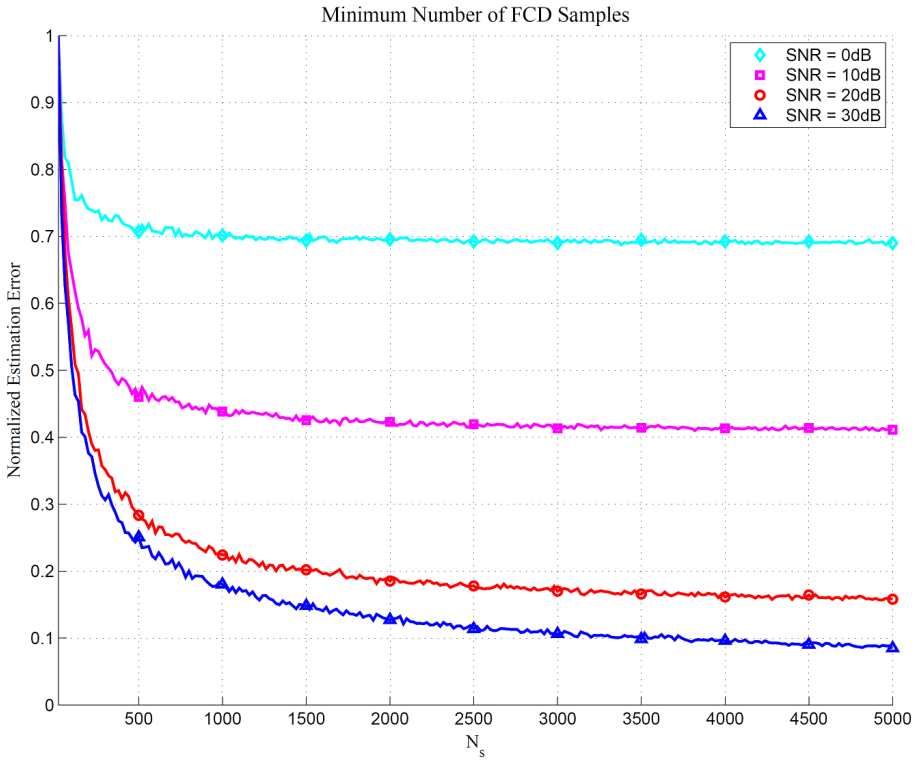

Figure 5.11: Traffic density signal recovery error. The number of required FCD samples to recover density signal varies under different signal-to-noise ratios. It depends on various parameters including noise power, road segment size, spatial distribution of points-of-interests and FCD samples.

mobile devices.

Chapter 6

Energy-Efficient State Recovery in Cyber Physical Systems

The compressive sensing theory has been intensively employed to develop new methods and applications. Energy conservation for every node in the network and overall energy consumption is one of the main design issues in such networks. In large-scale sensor networks, information is relatively sparse compared with the number of nodes. In such networks, the state recovery problem can be recast as a sparse signal recovery problem in the discrete spatial domain to be solved with a small number of linear measurements as an under-determined linear system by a ℓ_1-norm minimization program. A series of signal processing methods are applied to recover the current state of sensory data, e.g., temperature, pressure, force, flow, humidity, position or motion from noisy measurements. We propose an energy-efficient state recovery method for sensor networks. As a proof of concept, this method is employed to recover the air quality signal in an air quality monitoring system which is a dense measurement vector from wide geographically spread sensors. The signal contains the air quality indexes (AQI) of 866 weather stations that send their AQI indexes to a central base station through wireless channel. In a megascopic level the dynamic part (AC) of this signal may indicate the mobility in transportation system. On the other hand, the elements of this vector are linear combinations of chemical sensors with different modalities (e.g., ozone, sulfur dioxide, nitrogen dioxide, carbon monoxide). The signal itself is statistically similar to speed measurements in the transportation systems both time and frequency domains. Wide geographical distributions of the sensors in this system is the main reason to choose this type of date for evaluation of proposed energy efficient method.

To have a sparse representation of the signal first it is transformed to the frequency-

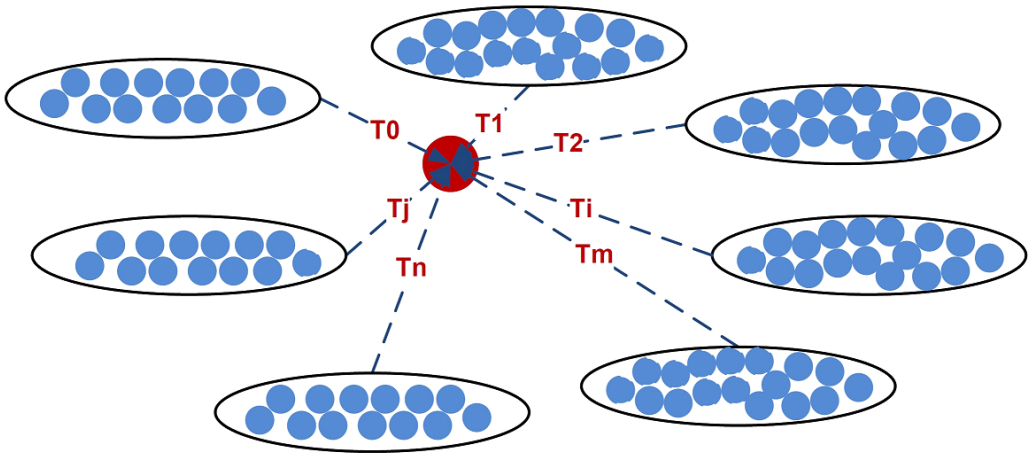

Figure 6.1: Energy-aware signal recovery. Measurement in the same modality at successive time instances are reported to a central station to be processed in reconstructing the original signal.

domain, then recovered and reconstructed from a small portion of coefficients. Different random matrices driven from Bernoulli and Gaussian distributions are investigated to find an energy-efficient sensing scheme for signal reconstruction. The results reveal more than sixty percent savings in power consumption with only a ten percent reconstruction and recovery error. The proposed method prolongs network lifetime with noticeable saving in deployment and maintenance costs, especially in large-scale sensor networks with slowly varying phenomena. Figure 6.1 is a sketch of problem that we tackle in this chapter.

6.1 Compressive-Sensing State Recovery

A sensor network is one of the essential parts in design and developing cyber-physical-systems. It generally consists of autonomous wireless nodes numbering from a few to several billions, e.g., cell-phones, used for area monitoring and environmental sensing by processing massive sensory data such as temperature, pressure, sound, video, vibration, force, flow, humidity, position and motion. These nodes are spatially distributed vary in size, power consumption, mobility, topology, communication scheme and bandwidth, interface, hardware, middle ware, and software. Energy conservation for every node and overall energy consumption in the network is one of the main design issues in such networks. It maintains the key role in the lifetime of the sensors and the network. If we consider the

outputs of deployed sensors as a time-domain signal components in the form of a vector or matrix, the theory of Compressive Sensing (CS) provides the opportunity to recover this signal from a small number of random samples compared with the Nyquist rate that states twice the maximum frequency sampling rate is required to recover a signal. Compressive sensing (a.k.a. compressed sensing, compressive sampling and sparse sampling) is a procedure for reconstructing a signal utilizing the prior sparse or compressible knowledge. The field has existed for several decades, but recently has been considered intensively again. As a consequence, it has been used to improve the performance and robustness of estimation methods through sparse information [32, 16, 99, 18, 8]. One of the primary applications of CS in wireless networks was introduced in [6] where the authors introduced and analyzed an energy-efficient estimation method. In [83], a mobile cooperative network is considered to build a map of the spatial variations in the environment. Another energy-efficient scheme by CS is proposed in [38] to achieve energy and bandwidth efficiency in the underwater sensor network. Some works have utilized CS to localize or track mobile nodes indoors via multi-modal data [21, 46, 107]. A comprehensive literature survey of CS-based theories, methods and applications is available on [57]. There is extensive research on static or dynamic state recovery methods that have recently been dominated by Bayesian techniques (i.e., Kalman filters, Particle filters, random finite sets) with multi-modal (i.e., radio, video, audio) data [24, 34].

In this work, we consider the state recovery problem for sensor networks through sparse and noisy measurements using compressive sensing theory. The state variable could be formulated as a vector of sensory data samples of position, velocity, acceleration, and temperate of nodes in a sensor network. We have applied the proposed method to collected sensory data in a Cyber Physical System.

The proposed state recovery method in this chapter utilizes CS theory to estimate the current state of n stationary sensors in the network via m measurements, while the number of measurements is smaller than the grid size ($m<n$). Here, the state recovery problem is recast as a sparse signal recovery problem in the discrete spatial domain. The method and solutions presented in this chapter can be easily adopted with all cyber physical systems with slight modifications. We have applied the CS theory to recover the current state of nodes in a monitoring system, in which the nodes send their current measurements to the monitoring center. Each node in such systems can range in size from a small box to the size of a building. In such networks, the cost of monitoring (i.e., operation time, power consumption) is a crucial design parameter. Consider a network with n stationary nodes over the time period T, distributed in a coverage area to report measurements over the same period of time. The state s_i^t of node i at time t_i can be defined as quadruple $s_i^t = (t_i, lon_i, lat_i, q_i^t)$. Figure 6.2 shows an example of such networks, which 866 nodes are

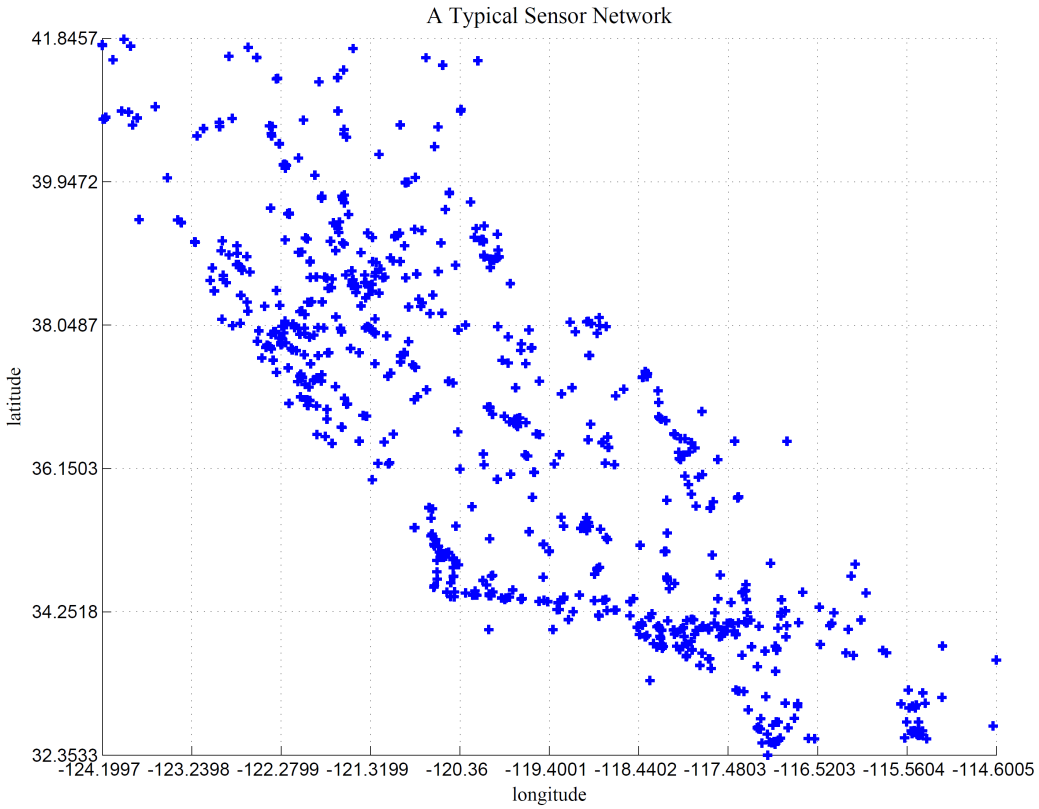

Figure 6.2: A typical cyber-physical system. Nodes report measurements to central station at time instance t. The state s_i^t of node i at time t_i can be defined as $s_i^t = (t_i, lon_i, lat_i, q_i^t)$. Nodes are distributed over longitudes [-124.1997,-114.6005] and latitudes [32.3533,41.8457].

distributed over California State to monitor daily pollution levels of different chemicals (e.g., ozone, sulfur dioxide, nitrogen dioxide, carbon monoxide) and particles (lower than 10 micrometers) in multiple cities and regions.

In most sensing systems, the information is sparse compared to the number of nodes (sparse state vector). In the case of a dense state vector a sparsifying operator Ψ can be applied first to the original state vector. In our approach, Discrete Cosine Transform has been used for sparsifying the time domain signal. The orthogonal bases such as DFT and DCT are ideal sparsifying bases with almost zero mutual coherencies. Figure 6.3 shows a 256-point DCT sparsifying matrix:

If we sketch the energy of the represented signal in frequency domain versus the number

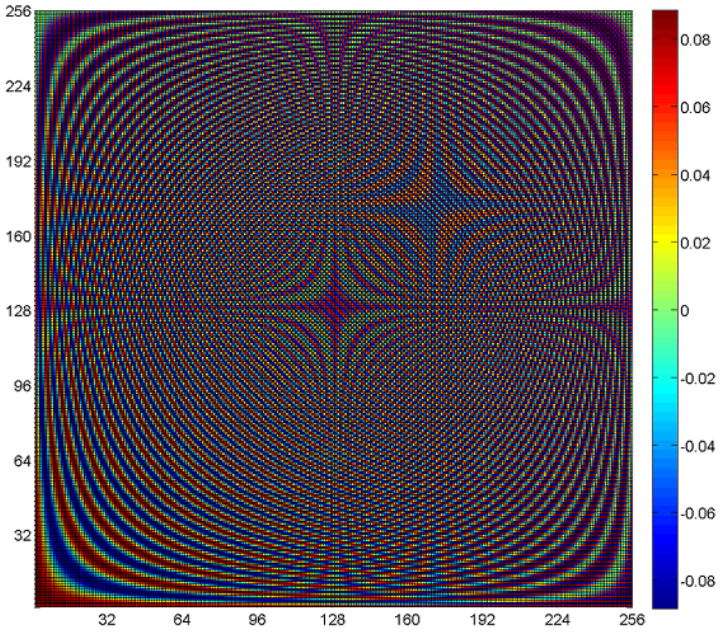

Figure 6.3: Discrete Cosine Transform as sparsifying basis. An ideal sparsifying matrix to convert dense observation signal to sparse one.

of components (Figure 6.4), we note that first 64 components (6% of signal length) of the signal in transform domain contain 99% of signal energy. At the same time we can recover the original signal with less than 10% recovery error with only 130 components (13% of signal length). Both cases reveal that our signal can be represented sparsely in transform domain ($k \simeq 0.13n \ll n$) with acceptable recovery error or degradation in signal quality. This error is accumulated with reconstruction error due to ℓ_1-minimization to solve under-determined system with a limited number of measurements. The sparsity-order k is dependent on original signal $x$and its transformation in the frequency domainand is usually low ($k < 0.2n$) for slowly varying signals. Highly changing signals in the time domain are more evenly distributed energies in frequency domain. We have used the DCT sparsifying basis as it has better energy distribution among the highest energy bases functions than DFT, with respect to MSE analysis.

There is no necessity to transfer the original signal to frequency domain if we have a k-sparse representation of it. In this case, we may reconstruct the signal in the time domain with a few measurements. The performance is affected by the measurement matrix (A) and measurement noise (ϵ).

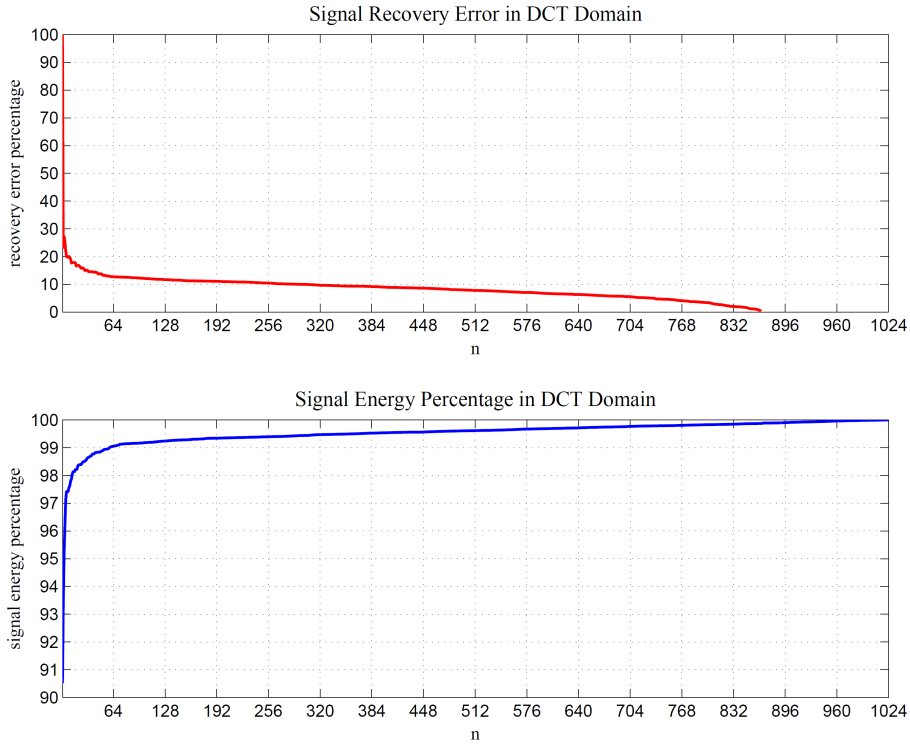

Figure 6.4: Signal energy and recovery error in DCT domain. The first 64 components of the signal in transform domain contain 99% of signal energy. The MSE analysis reveals that original signal can be recovered with less than 10% recovery error with only 130 components.

Figure 6.5 depicts the original signal x and sparsified matrix z. Vector x contains 866 measurements from all available nodes. To have a sparse representation z of signal x, first we apply a 1024-point DCT to transform it to the frequency domain.

Figure 6.6 shows the CS-based recovery in the frequency domain through sparse representation of the signal. Original signal x in length of 866 is transferred to the frequency domain to get the sparse representation z. First 64 components ($k = 64 \ll 1024$) in the transform domain are selected for sparse reconstruction from 130 measurements ($m = 130$). The reconstructed signal in the frequency domain is then recovered to the time domain.

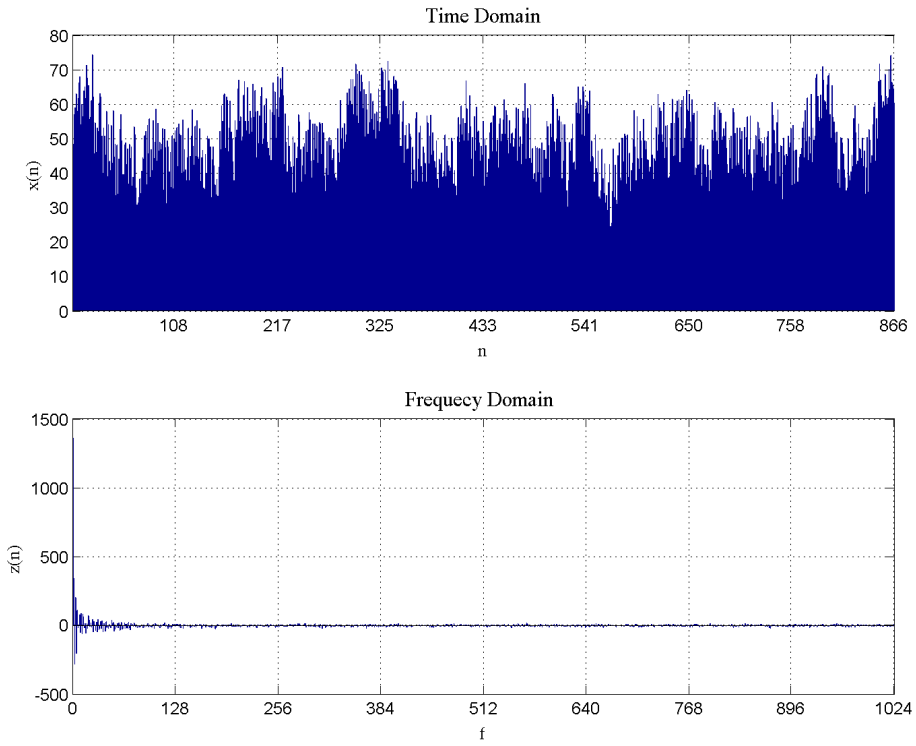

Figure 6.5: Time-domain measurements and its sparse representation in DCT domain. The collected measurements from 866 nodes are sent to monitoring center. Discrete Cosine Transform z of original dense signal x can be represented in sparse format.

6.2 Performance and Experimental Results

The performance (estimation error) of the proposed method depends on several parameters, i.e., sparsity order k, observation order (number-of-measurements) m, sparsifying basis Ψ, sensing matrix Φ or measurement matrix A, as well as the measurement noise ϵ. All estimations errors are calculated based on the normalized mean squared error which is defined as (6.1) for signal s and the estimated one $\hat{s}$.

$$e_n = \frac{\| s - \hat{s} \|_2}{\| s \|_2} \tag{6.1}$$

Figure 6.7 sketches the variation on normalized mean-squared-error e_n versus the number-of-measurements m under different sparsity orders k for the ternary measurement matrix

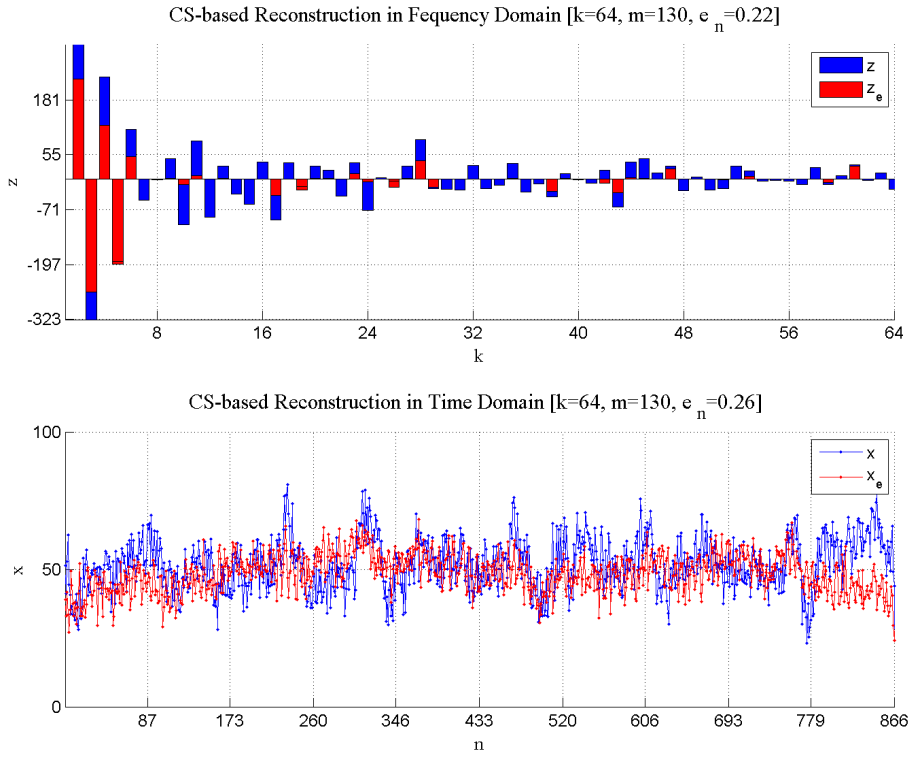

Figure 6.6: CS-based signal reconstruction. Original signal x is transformed to DCT domain. An ideal sparsifying matrix is applied to original dense observation signal x to have sparse representation z. The transformed signal z is recovered from its first $k = 64$ highest energy components through only $m = 130$ measurements. The normalized estimation error for $\hat{x}$ and $\hat{z}$ are $e_n = 0.13$ and $e_n = 0.19$ respectively.

shown (Figure 2.4), which is in size of 130×1024 for our case. The results are averaged over the 1000 different measurements matrices. To recover the signal in the transform domain with low recovery error, higher sparsity orders are required. On the other hand, higher sparsity orders need a higher number-of-measurements in the network to achieve better performance in signal reconstruction through ℓ_1-minimization.

Figure 6.8 shows the compressed reconstruction performance evaluation for the same signal in the time domain. The results are averaged over 1500 iterations with different sparsity orders k and number-of-measurements m.

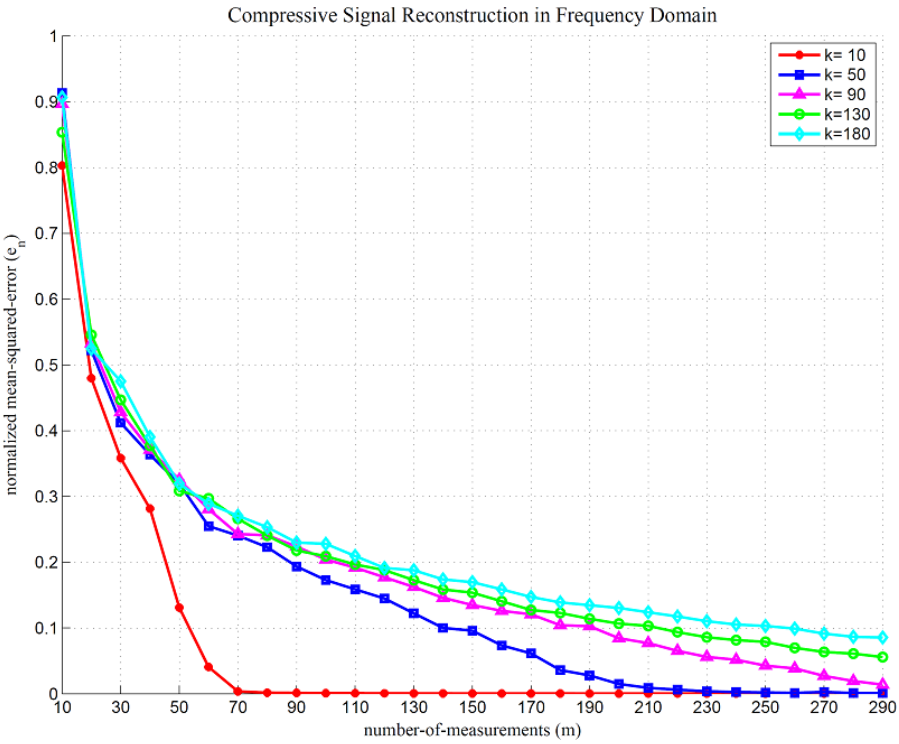

Figure 6.7: Compressive signal reconstruction in frequency domain for ternary sensing matrix with Bernoulli distribution in Figure 2.4. To recover signal in transform domain with less recovery error, higher sparsity orders are required. On the other hand, higher sparsity order leads to higher number of measurements in the network to achieve better performance in signal reconstruction through ℓ_1-minimization.

6.2.1 Comparison with Kalman Filtering

The Kalman filter addresses the general state estimation problem for a state variable $x \in R^n$ of a discrete-time process with a linear stochastic difference equation:

$$x_k = A_{n\times n} \times x_{k-1} + B_{n\times 1} \times u_{k-1} + w_{k-1} \tag{6.2}$$

From measurements $y \in R^m$ that is:

$$y_k = H_{m\times n} \times x_k + v_k \tag{6.3}$$

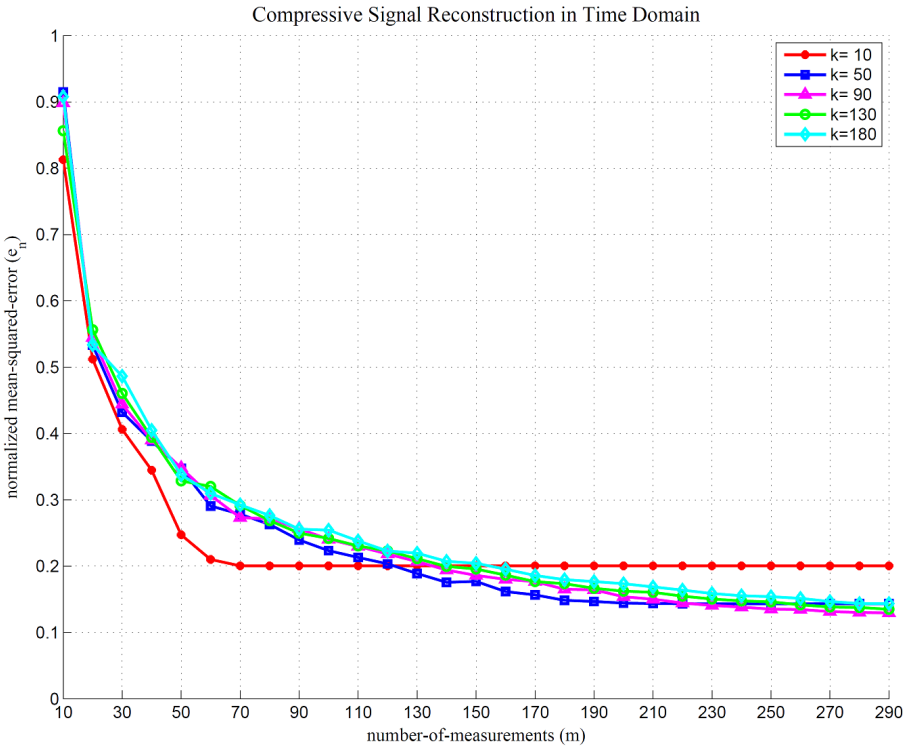

Figure 6.8: Compressive signal reconstruction in time domain for ternary sensing matrix with Bernoulli distribution in Figure 2.4. We have extra degradation due to sparse selection of signal in frequency domain.

The random variables $w_k \sim N(0, Q)$ and $v_k \sim N(0, R)$ represent the process and the measurement noise respectively, that are independent and have normal distributions with Q and R noise covariance matrices. They are time-dependent but for simplicity, we usually assume them to be constant [119].

If we compare measurement equation (6.3) with the CS main equation (4.1), we can rephrase our compressive sensing problem as a state estimation one to be solved with the classic Kalman filtering approaches. As the measurements for all stations are not available in consecutive time samples (t_i), the input signal state x is only updated based on the availability of its corresponding components (non-zero values in measurement matrix A).

The CS-based schemes can be easily implemented by a simple sequencer in nodes. Every column in measurement matrix defines the sequence of measurements (on/off) for each node in the network. The interval of measurements (i.e., rows in measurements matrix)

can be set regarding to the signal dynamics. Obviously, shorter measurement intervals are required in highly changing environments.

$$\begin{matrix} t_1 \\ \vdots \\ t_m \end{matrix} \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix}$$

In practice, the process noise covariance Q and the measurement noise covariance R matrices might change with each time step or measurement; however, here we assume they are constant. Figure 6.9 shows the comparison results for the compressive sensing (CS) and the Kalman filtering (KF) methods. The sparsity order $k = 130$ is set for a state vector in size of the original signal vector. The results are compared for different a number of measurements. Estimation errors converge to the same values whenever the number of available measurements increases.

One of the main interesting comparisons between these different state recovery methods is the computational time required for each. Figure 6.10 plots the computational time of the methods versus the variable number of measurements. It is noticeable that the processing time is not monotonically increasing for the compressive sensing as we solve an optimization problem here with variable convergence time. The Kalman filter has a linear increase in processing time compared to the number of measurements.

6.2.2 Energy Saving

One of the main goals in utilizing the CS theory is signal reconstruction in a cost-efficient way. In the wireless sensor network, cost is related to different parameters, i.e., maintenance, deployment, communication and energy costs. Generally, the sensor cost is considered as a weight vector in size of n that must be optimized in a deterministic or probabilistic way. In cyber physical systems, communications between sensors and the central station are usually wireless. In wireless communicationswhen the distance between transmitter (sensor) and receiver (base station) is much larger than the wavelength of the transmitted wavethe received power is given by:

$$p_r = p_t g_t g_r \left(\frac{c}{4\pi d f_c} \right)^2 \tag{6.4}$$

where,

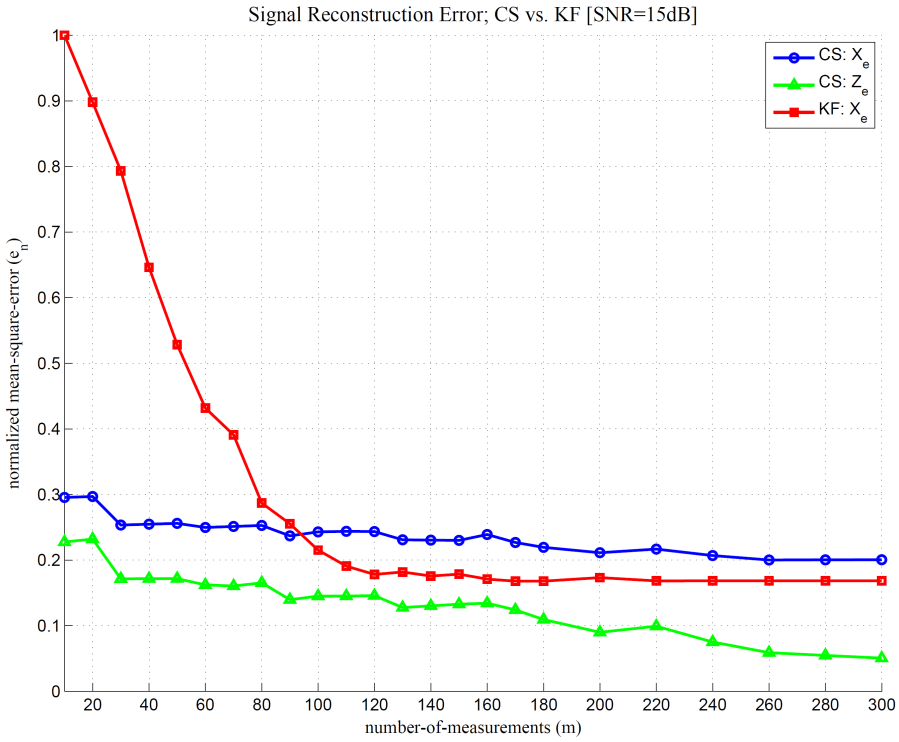

Figure 6.9: Signal reconstruction process; Compressive Sensing vs. Kalman Filtering. Available measurements at consecutive time samples are used for state vector estimation. As the number of measurements increase, the estimation error converges to its limit value.

- p_r, p_t: receiver/transmitter power
- g_r, g_t: receiver/transmitter gain
- d: distance between transmitter and receiver
- c: light speed
- f_c: carrier frequency

Simply, the transmitted power is proportional to the square of the distance between the transmitter and receiver ($p_t \propto d^2$) [74].

If we suppose transmitter power (p_t) as the sensor cost, then in an isometric environment, the weight vector $\vec{\omega} = (\omega_1,\ \omega_2,\ \ldots,\ \omega_n)$ can be rephrased as $\vec{\omega} = (d_1^2,\ d_2^2,\ \ldots,\ d_n^2)$, where d_i is the distance between the sensor and base station.

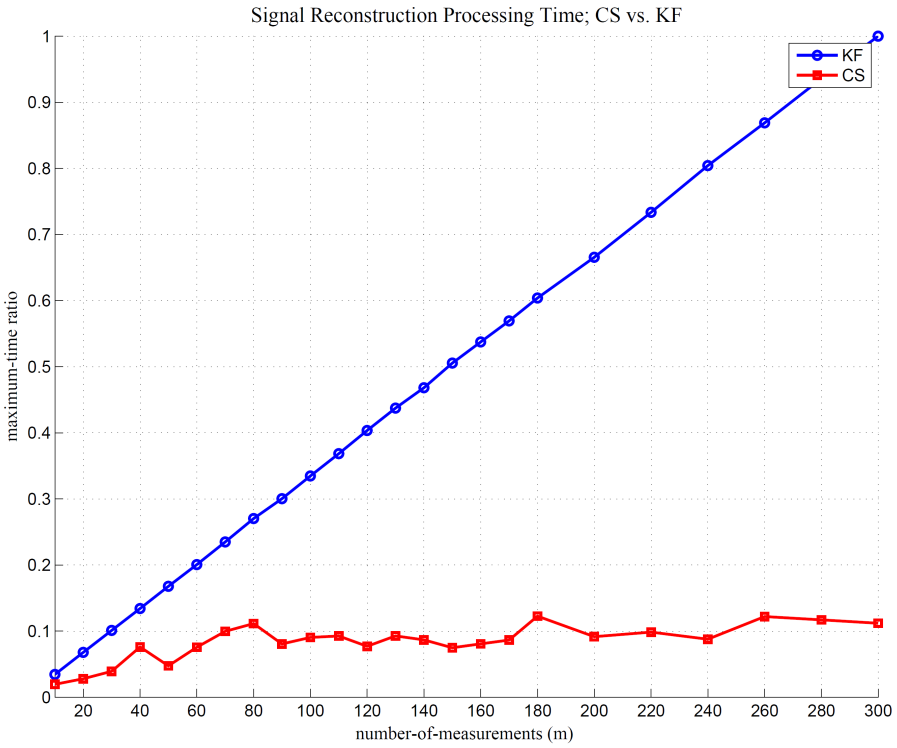

Figure 6.10: Signal reconstruction time. The plot shows relative processing time for CS and KF methods for a state vector in size of 866 from $m = 300$ measurements. It is non-monotonic for CS, as an optimization problem is solved with variable convergence time.

One can save up to 65% in power consumption with a reasonable sparsity order $k = 130$. To maintain an acceptable signal reconstruction error using compressive sensing by saving on power consumption, one can design different a measurement matrix for sensing. The ultimate goal in designing such matrices is having the minimum number-of-measurements to maintain a predefined signal reconstruction error.

Figure 6.11 plots the relative power consumption versus the number of measurements for Non-CS (linear) and CS (i.e., Bernoulli and Gaussian) approaches. Within CS models, Gaussian measurements matrices reconstruct the signal with fewer measurements in the same sparsity order.

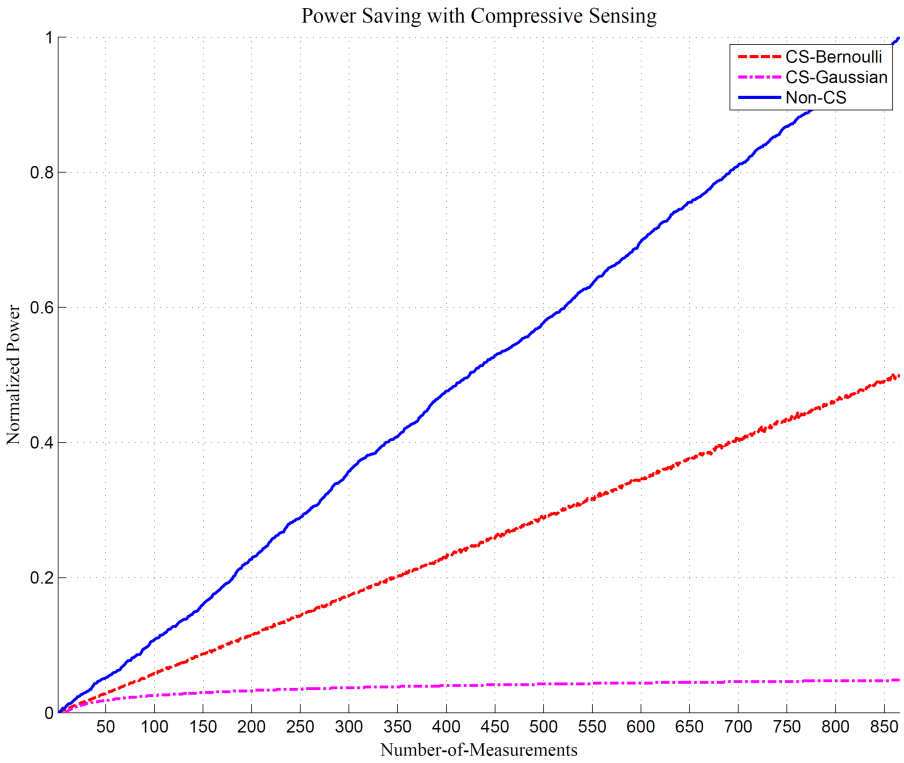

Figure 6.11: Power consumption using CS and non-CS methods. Normalized power consumption is sketched vs. to the number of deployed sensors in Non-CS, CS-Binomial, and CS-Gaussian scenarios. In this results, all sensors have the same operational cost (uniform cost vector).

6.3 Chapter Summary

In this chapter, we proposed an energy efficient state recovery method for large-scale sensor networks based on the compressive sensing theory. A sparse representation of sensory data was proposed to reconstruct it later from a few noisy measurements compared to the number of nodes. The problem was recast as an under-determined linear system and the sparse solution was provided by ℓ_1-minimization. In case of a dense signal in the time domain, we may transform the signal to frequency domain to have a sparse representation, so having compressible sensory data is an underlying condition for the proposed method. We investigated different sensing matrices to have the most energy-efficient state recovery scheme. The number of components for signal recovery in the transform domain must meet

the signal recovery error requirements. The performance of the method was assessed for power consumption and signal reconstruction error through mean-squared-error analysis. The proposed method can be adopted for large-scale sensor networks with slowly varying signals that are highly compressible in the transform domain. Applying this method to massive sensory data in wireless sensor networks may improve the resource utilization in many demanding applications.

Chapter 7

Conclusion

The focus of this thesis has been the design and implementation of various statistical filters to process multimodal sensory data in CPS for mobility modeling. While the use of such filters in the context of data fusion and modeling is relatively common-place, their use in CPS for mobility modeling is relatively new, and represents an attractive option for multimodal sensory systems. Several methods for data fusion and state recovery from sparse and inconsistent measurements are presented. Finally, an energy efficient state recovery method is presented. Among these statistical filters, compressive sensing represents an excellent candidate for processing multimodal sensory data in energy aware applications.

7.1 Summary of Contributions

In this thesis the benefits of utilizing statistical filtering methods in mobility modeling have been studied. The thesis attempted to show the efficiency of these techniques in processing massive streams of sensory data in cyber-physical systems. This thesis makes several contributions to the field of data fusion and mobility modeling:

- An adaptive particle weighting for multimodal data fusion presented. Since the introduction of optimal estimation for non-linear, and non-Gaussian state-space models, particle filters have become a popular class of algorithms to solve such problems numerically from recursive observations. Particle filters provide an efficient and flexible way to approximate nonlinear functions. The efficiency and accuracy of the particle filters depend mainly on two key factors: the number of particles used and the

propagation function to re-allocate these particles at each iteration. We proposed a novel multimodal data fusion approach that applied for localization of mobile users in cellular networks. First, a series of vision-based algorithms applied to derive a preliminary mobile user position from monocular vision to augment with estimated location in a cellular network. These algorithms have been used for feature extraction, feature matching, perspective-n-point depth inversing. In the next step, particle filters employed to fuse bimodal radio-visual position estimations via signal strength and monocular vision measurements. This approach improves the accuracy of signal strength-based localization methods by augmenting it with visual information. We show that fusing frequent inaccurate measurements with low-frequent accurate ones may increase the performance of localization significantly. This fusion can take place at different processing stages of mobility modeling. The main limitation of the proposed method is the fact it requires an up to date video capture of the environment. Although, the experiments show that the proposed method is robust against environmental changes, it may not be suitable for highly changing environments, for example, in environments with parked cars, moving people, trees, and buildings that their visual characteristics (i.e., shape, color, brightness) vary across video instances. Occlusion can have a significant impact on the system performance.

- A parametric kernel optimization platform for mobility modeling developed. Traffic estimation and prediction services play a key role in intelligent transportation systems. The performance of such systems is heavily dependent on the availability of traffic flow information. The ability of the system to analyze the different types of floating car data also impacts estimates. We examined the problem of traffic density modeling in transportation networks and proposed a kernel-based density optimization technique to resolve this problem. The weights of the Gaussian mixtures optimized through minimization of the Cramer-von Mises distance between the localized cumulative distributions of the derived position samples and the points-of-interest in traffic modeling. The collected position samples of floating car data was modeled with Generalized Gaussian Density. Then, the distance between the localized cumulative distributions of the probability densities centered at floating cars positions and points-of-interests was computed. This distance was minimized through quadratic programming. From these optimized weights, the probability distribution function around new points-of-interests can be interpolated or estimated. The approach was applied to collected GPS data to evaluate its performance. The performance of the system depends on many parameters including the bandwidth of kernels as well as the trade-off factor on regularization that can be defined in an adaptive way for future work. Having a spatial-temporal model also helps us in defining the kernels'

bandwidth as a Priori in the optimization process. The proposed method can be easily adopted to model the traffic evolution in urban areas under different traffic conditions. It can be also employed for congestion detection or movement pattern analysis indoors using the collected position samples from hand-held mobile devices.

- An Energy-efficient state recovery proposed. The compressive sensing theory has intensively inspired the development of the new methods and applications. Energy conservation for every node in the network and the overall power consumption are key design issues in such networks. In large-scale sensor networks, information is relatively sparse compared with the number of nodes. In such networks, the state recovery problem can be stated as a sparse signal recovery problem in the discrete spatial domain. It can be solved with a small number of linear measurements as an under-determined linear system by a ℓ_1-norm minimization program. We proposed an energy efficient state recovery method for large-scale sensor networks based on the compressive sensing theory. A sparse representation of sensory data was proposed to reconstruct it later from a few noisy measurements compared to the number of nodes. The problem was recast as an under-determined linear system and the sparse solution was provided by ℓ_1-minimization. In case of a dense signal in the time domain, we may transform the signal to frequency domain to have a sparse representation, so having compressible sensory data is an underlying condition for the proposed method. We investigated different sensing matrices to have the most energy-efficient state recovery scheme. The number of components for signal recovery in the transform domain must meet the signal recovery error requirements. The performance of the method was assessed for power consumption and signal reconstruction error through mean-squared-error analysis. The proposed method can be adopted for large-scale sensor networks with slowly varying signals that are highly compressible in the transform domain. Applying this method to massive sensory data in wireless sensor networks may improve the resource utilization in many demanding applications.

7.2 Future Work

The integration of multimodal sensory data is critical for over-coming any performance issues related to inaccurate measurements in CPS. This data integration (i.e. fusion) can be implemented in several categories or levels such as data alignment, entity assessment, situation assessment, impact assessment, process refinement, and user refinement. We proposed a method to fuse sensory date at the lowest level (i.e. data alignment). Unfortunately, at

this level, inaccuracy in data measurements may impact the system's performance significantly. One can improve the proposed method by investigating the sensor fusion in higher levels (e.g. entity or user). This type of improvement would be a significant enhancement for this type of system. Other filtering methods than particle filters can be employed to benefit from their unique features.

There are several problems that can be explored to enhance the performance of the proposed CS-based methods and applications (i.e. state recovery and positioning). Besides the main sensor in an application, most of the devices in CPS are now equipped with extra sensors. The information generated from these sensors can be integrated with the main sensing measurements. Recent mobile sensory nodes are equipped with high-performance processors. This allows the mobile nodes to be used in processing the sensory data and implementation of proposed methods distributively. Thus, the proposed methods should be extended to fuse the information obtained from extra sensors to improve the system's performance.

In addition to using the CS theory to reduce the number of measurements, empirical models or tracking methods may be used in conjunction to interpolate data when no measurement is available. An empirical model can also be used to design adaptive measurement matrices for a specific application.

The performance of the proposed methods is heavily dependent on quality of measurements, thus data acquisition plays a key role in all comparisons among similar methods on the same applications. As a result, only a very limited amount of data sets are available which can be used to test algorithms. Even fewer experiments have been carried out specifically for CS-based methods. Experiments must be planned for the near future to provide benchmarks for compressive sensing algorithms and results from these experiments can drive future research.

References

[1] *Wireless Ad-Hoc and Sensor Networks*, chapter 1: A Survey of Mobility Models. Kluwer Academic Publishers, 2008.

[2] S. Aeron, V. Saligrama, and M. Zhao. Information theoretic bounds for compressed sensing. *IEEE Transactions on Information Theory*, 56(10):5111–5130, 2010.

[3] H. Alkhatib, I. Neumann, H. Neuner, and H. Kutterer. Comparison of sequential monte carlo filtering with kalman filtering for nonlinear state estimation. International Conference on Machine Control and Guidance, 2008.

[4] A. Amini and F. Marvasti. deterministic construction of binary, bipolar, and ternary compressed sensing matrices. *IEEE Transactions on Information Theory*, 57(4):2360–2370, 2011.

[5] T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping: PART II. *IEEE Robotics and Automation Magazine*, 13(3):108–117, September 2006.

[6] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak. Compressive wireless sensing. *in Proceedings of The 5Th International Conference on Information Processing in Sensor Networks (IPSN '2006)*, pages 134–142, 2006.

[7] R. G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118–121, July 2007.

[8] R. G. Baraniuk, E. J. Candes, R. Nowak, and M. Vetterli. Compressive sampling. *Ieee Signal Processing Magazine*, 25(2):12–13, 2008.

[9] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.

[10] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[11] A. Basu, H. Shioya, and C. Park. *Statistical Inference: The Minimum Distance Approach*. 2011.

[12] A. Bensky. *Wireless positioning technologies and applications*. Artech House, 2008.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. 2004.

[14] C. Brown. Tutorial on filtering, restoration, and state estimation. Technical Report 534, University of Rochester, The University of Rochester, Computer Science Department, Rochester, New York, 14627, 1995.

[15] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad-hoc network research. *Wireless Communications and Mobile Computing*, 2(5):483–502, September 2002.

[16] E. J. Candes. Compressive sampling. *International Congress of Mathematicians*, 3(2):1433–1452, 2006.

[17] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2010.

[18] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[19] E. J. Candes and M. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[20] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[21] V. Cevher, P. Boufounos, R. G. Baraniuk, A. C. Gilbert, and M. J. Strauss. Near-optimal bayesian localization via incoherence and sparsity. *2009 International Conference on Information Processing in Sensor Networks (IPSN'2009)*, pages 205–216, 2009.

[22] R. Chen and J. Liu. Mixture kalman filters. *journal Royal Statistical Society*, 62:493–508, 2000.

[23] S. Chen, D. L. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[24] Z. Chen. Bayesian filtering from kalman filters to particle filters and beyond. Technical report, Natural Sciences and Engineering Research Council of Canada, 2002.

[25] D. Cobzas, H. Zhang, and M. Jagersand. Image-based localization with depth-enhanced image map. *International Conference on Robotics and Automation*, 2003.

[26] C. F. Daganzo. Requiem for second-order fluid approximations of traffic flow. *Transportation Research B*, 29B(4):277–286, 1995.

[27] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. *9th International Conference on Computer Vision*, 2003.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38, 1977.

[29] R. A. DeVore. Deterministic constructions of compressed sensing matrices. *journal of Complexity*, 23:918–925, 2007.

[30] M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. *IEEE Transactions on Image Processing*, 11(2):146–158, 2002.

[31] D. L. Donoho. for most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution. *Technical Report, Communication and Pure Applied Math, Department of Statistics, Stanford University*, 2004.

[32] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[33] D. L. Donoho, H. Kakavand, and J. Mammen. The simplest solution to an underdetermined system of linear equations. *2006 IEEE International Symposium on Information Theory*, pages 1924–1928, 2006.

[34] A. Doucet, N. de Freitas, , and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer-Verlag, 2001.

[35] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *journal Royal Statistical Society*, 10(3):197–208, 2000.

[36] C. R. Drane. *Positioning systems in intelligent transportation systems*. Artech House, 1998.

[37] H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping: PART I. *IEEE Robotics and Automation Magazine*, 13(2):99–110, June 2006.

[38] F. Fazel, M. Fazel, and M. Stojanovic. Random access compressed sensing for energy-efficient underwater sensor networks. *IEEE Journal on Selected Areas in Communications*, 29(8):1660–1670, 2011.

[39] M. Fazel, E. J. Candes, B. Recht, and P. Parrilo. Compressed sensing and robust recovery of low rank matrices. *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047, 2008.

[40] C. Feng, S. Valaee, and T. Zhenhui. Multiple target localization using compressive sensing. *2009 IEEE Global Telecommunications Conference (GLOBECOM'2009)*, pages 1–6, 2009.

[41] J. Folkesson, P. Jensfelt, and H. I. Christensen. Vision SLAM in the measurement subspace. In *Proceeding of the IEEE International Conference on Robotics and Automation*, 2005.

[42] D. A. forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Intelligent robotics and autonomous agent series. Prentice Hall, 1st edition, August 2002.

[43] S. Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Applied and Computational Harmonic Analysis*, 29(1):97–103, 2010.

[44] T. Goedeme, T. Tuytelaars, L. V. Gool, G. Vanacker, and M. Nuttin. Feature based omnidirectional sparse visual following. *International Conference on Intelligent Robots and Systems*, pages 1003–1008, 2005.

[45] D. O. Gorodnichy and W. W. Armstrong. Single camera stereo for mobile robot world exploration. In *Proceeding of Vision Interface Conf. VI*, 1999.

[46] A. Griffin and P. Tsakalides. Compressed sensing of audio signals using multiple sensors. *in Proceeding of 16Th European Signal Processing Conference (EUSIPCO'08)*, 2008.

[47] U. D. Hanebeck and V. Klumpp. Localized cumulative distributions and a multivariate generalization of the cramer-von mises distance. *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 33–39, 2008.

[48] R. Haralick, C. Lee, K. Ottenberg, and M. Nolle. Review and analysis of solutions of the three point perspective pose estimation problem. *International journal of Computer Vision*, 13(3):331–356, 1994.

[49] C. Harris and M. Stephens. A combined corner and edge detector. *Alvey Vision Conference*, pages 147–151, 1988.

[50] M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. Wiley, 2nd edition, 1996.

[51] G. Heine. *GSM Networks: Protocols, Terminology and Implementation (Mobile Communications Library)*. Artech House Publishers, 1st edition, 1998.

[52] J. C. Herrera, D. Work, X. Ban, R. Herring, Q. Jacobson, and A. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: the mobile centuryfield experiment. *Transportation Research C*, 18:568–583, 2010.

[53] P. Ho, Wang Y., Hou F., and Shen S. A study on vertical handoff for integrated wlan and wwan with micro-mobility prediction. In *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*, pages 1 –11, oct. 2006.

[54] S. P. Hoogendoorn and P. H. L. Bovy. State-of-the-art of vehicular traffic modeling. *journal of System and Control Engineering*, 215:283–303, 2001.

[55] C. Hsu and F. Lian. A case study on highway flow model using 2-d gaussian mixture modeling. *2007 IEEE Intelligent Transportation Systems Conference*, pages 790–794, 2007.

[56] W. J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *IEEE/ACM Transactions on Networking*, 17(5):1564–1577, October 2009.

[57] Rice University Compressive Sensing Resources. [Online]. Available: http://dsp.rice.edu/cs/.

[58] W. Y. Jeong and K. M. Lee. Visual SLAM with line and corner features. In *Proceeding of the IEEE/RSJ International Conference on Intelligent Robotics and System*, Beijing, China, October 2006.

[59] B. Kashin and V. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82:748–755, 2007.

[60] M. Khalaf-Allah. Nonparametric bayesian filtering for location estimation position tracking and global localization of mobile terminals in outdoor wireless environments. *EURASIP journal on Advances in Signal Processing*, (317252):14, 2008.

[61] K. Kidono, J. Miura, and Y. Shirai. Autonomous visual navigation of a mobile robot using a human-guided experience. *Robotics and Autonomous Systems*, 40(2-3):124–132, 2002.

[62] D. Koks and S. Challa. An introduction to bayesian and dempster-shafer data fusion. Technical Report DSTO-TR-1436, Department of Defence, DSTO Systems Sciences Laboratory, P.O. Box 1500, Edinburgh, SA 5111, Australia, November 2005.

[63] P. Krauthausen and U. D. Hanebeck. Regularized non-parametric multivariate density and conditional density estimation. *2010 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 180–186, 2010.

[64] P. Krauthausen, M. Roschani, and U. D. Hanebeck. Incorporating prior knowledge into nonparametric conditional density estimation. *American Control Conference (ACC'2011)*.

[65] R. D. Kuhne and P. Michalopoulos. Traffic flow monograph: Chapter 5. *FHWA, U.S. Department of Transportation*, pages 21–25, 1998.

[66] M. Lai. On sparse solutions of underdetermined linear systems. *Department of Mathematics the University of Georgia Athens*, 2009.

[67] M. Li, B. Hong, Z. Cai, and R. Luo. Novel rao-blackwellized particle filter for mobile robot SLAM using monocular vision. *International Journal of Intelligent Technology*, 1(1), 2006.

[68] Z. Li, F. Wu, and J. Wright. on the systematic measurement matrix for compressed sensing in the presence of gross errors. *2010 Data Compression Conference*, pages 356–365, 2010.

[69] M. J. Lighthill and G. B. Whitham. on kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178):317–345, 1955.

[70] D. Lin and R. Juang. Mobile location estimation and tracking for gsm systems. *IEEE Transactions on Vehicular Technology*, 54(4):1447–1454, July 2005.

[71] D. B. Lin and R. T. Juang. Mobile location estimation based on differences of signal attenuations for gsm systems. *IEEE Transactions on Vehicular Technology*, 54(4):1447–1454, 2005.

[72] G. Liu, A. S. Lyrintzis, and P. B. Michalopoulos. Improved high-order model for freeway traffic flow. *Transportation Research Record*, 1644:37–46, 1998.

[73] D. M. Malakoff. Bayes offers new way to make sense of numbers. *Science*, 286:1460–1464, 1999.

[74] J. W. Mark and W. Zhuang. *Wireless Communications and Networking*. 2003.

[75] Y. Matsumoto, M. Inaba, and H. Inoue. Visual navigation using view-sequenced route representation. In *International Conference on Robotics and Automation*, pages 83–88, 1996.

[76] A. D. May. *Traffic Flow Fundamentals*. 1990.

[77] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. 1996.

[78] G. McLachlan and D. Peel. *Finite Mixture Models*. 2000.

[79] N. Metropolis and S. Ulam. The monte carlo method. *American Statistics Association*, 44:335–341, 1949.

[80] P. G. Michalopoulos, P. Yi, and A. S. Lyrintzis. Development of an improved high-order continuum traffic flow model. *Transportation Research Record*, 1365:125–132, 1992.

[81] L. Mihaylova, D. Angelova, S. Honary, D. R. Bull, C. N. Canagarajah, and B. Ristic. Mobility tracking in cellular networks using particle filtering. *IEEE Transactions on Wireless Communications*, 6(10):3589–3599, October 2007.

[82] H. Mohimani, M. Babaie-Zadeh, and C. Jutten. A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm. *IEEE Transactions on Signal Processing*, 57(1):289–301, 2009.

[83] Y. Mostofi and P. Sen. Compressive cooperative sensing and mapping in mobile networks. *IEEE Transactions on Mobile Computing*, 10(12):1769–1784, 2011.

[84] E. F. Nakamura, A. F. Loureiro, and A. C. Frery. Information fusion for wireless sensor networks: Methods, models, and classifications. *ACM Computing Surveys*, 39(3), August 2007.

[85] X. D. Nguyen, B. J. You, and S. R. Oh. A simple framework for indoor monocular SLAM. *International journal of Control Automation, and Systems*, 6(1):62–75, February 2008.

[86] D. Nister. An efficient solution to the inverse-point relative pose problem. *Conference on Computer Vision and Pattern Recognition*, pages 147–151, 2003.

[87] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara. Accurate gsm indoor localization. *Ubicomp*, pages 141–158, 2005.

[88] M. Papageorgiou. Some remarks on macroscopic traffic flow modeling. *Transportation Research A*, 32(5):323–329, 1998.

[89] M. Papageorgiou, J. M. Blosseville, and H. Hadj-Salem. Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in paris: PART II: Coordinated on-ramp metering. *Transportation Research Part A: General*, 24(5):361–370, 1990.

[90] M. Papageorgiou, J.M. Blosseville, and H. Hadj-Salem. Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in paris: PART I: Modelling. *Transportation Research Part A: General*, 24(5):345–359, 1990.

[91] N. Patwari, A. O. III Hero, M. Perkins, N. S. Correal, and R. J. O'Dea. Relative location estimation in wireless sensor networks. *IEEE Transactions on Signal Processing*, 51(8):2137–2148, 2003.

[92] H. J. Payne. Models of freeway traffic and control. *Mathematical Models of Public Systems, edited by G.A. Bekey (Simulation Council, La Jolla, CA)*, 1:51–61, 1971.

[93] M. Pupilli and A. Calway. Real-time camera tracking using a particle filter. *Proceeding of British Machine Vision Conference*, 2005.

[94] X. Qin. Traffic flow modeling with real-time data for on-line network traffic estimation and prediction. *Ph.D. dissertation, Faculty of the Graduate School, University of Maryland, College Park, MD*, 2006.

[95] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory*, 54(5):2210–2219, 2008.

[96] H. Raza and P. Ioannou. Vehicle following control design for automated highway systems. *IEEE Control Systems Magazine*, 16(6):43–60, 1996.

[97] H. Ren and M. Q. Meng. Power adaptive localization algorithm for wireless sensor networks using particle filter. *IEEE Transactions on Vehicular Technology*, 58(5):2498–2508, 2009.

[98] B. Ristic. *Beyond the Kalman Filter: Particle Filters for Tracking Application*. Artech House, 2004.

[99] J. Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, March 2008.

[100] E. Royer, M. Lhuillier, M. Dhome, and J. M. Lavest. Monocular vision for mobile robot localization and autonomous navigation. *International journal of Computer Vision*, 74(3):237–260, 2007.

[101] B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. 2002.

[102] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International journal of Robotic Research*, 21(8):735–760, 2002.

[103] C. Sommer and F. Dressler. Progressing toward realistic mobility models in vanet simulations. *IEEE Communications Magazine*, 46(11):132–137, November 2008.

[104] M. Sun. Multimodal signal fusion. pages 1556–1557. ICME 2009, IEEE International Conference on Multimedia and Expo, June-July 2009.

[105] A. Tabibiazar and O. Basir. Bimodal localization in cellular networks utilizing particle filters. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1 –6, july 2010.

[106] A. Tabibiazar and O. Basir. Radio-visual signal fusion for localization in cellular networks. In *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010 IEEE Conference on*, pages 150 –155, sept. 2010.

[107] A. Tabibiazar and O. Basir. Compressive sensing indoor localization. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 1986 –1991, oct. 2011.

[108] A. Tabibiazar and O. Basir. Kernel-based modeling and optimization for density estimation in transportation systems using floating car data. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 576 –581, oct. 2011.

[109] A. Tabibiazar and O. Basir. Kernel-based optimization for traffic density estimation in its. In *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pages 1 –5, sept. 2011.

[110] A. Tabibiazar and O. Basir. Energy-efficient compressive state recovery from sparsely noisy measurements. *Instrumentation and Measurement, IEEE Transactions on*, 61(9):2392 –2400, sept. 2012.

[111] H. Tanizaki. *Nonlinear Filters*. Springer-Verlag, 2nd edition, 1996.

[112] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. Intelligent Robotics and Autonomous Agent Series. MIT Press, 2006.

[113] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, December 2007.

[114] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

[115] L. Vacchetti, V. Lepetit, and P. Fua. Stable 3d tracking in real-time using integrated context information. In *Conference on Computer Vision and Pattern Recognition*, Madison, WI.

[116] M. K. Varanasi and B. Aazhang. Parametric generalized gaussian density estimation. *journal of Acoustical Society of America*, 86:1404–1415, 1989.

[117] S. Vavasis. Derivation of compressive sensing theorems for the spherical section property. *University of Waterloo, CO769 lecture notes*, 2009.

[118] V. Vivekanandan and V. Wong. Concentric anchor beacon localization algorithm for wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 56(5):2733–2744, 2007.

[119] G. Welch and G. Bishop. An introduction to the kalman filter. *1995*.

[120] G. B. Whitham. *Linear and Nonlinear Waves (Pure and Applied Mathematics)*. 1974.

[121] Z. Yang and Y. Liu. Quality of trilateration: Confidence-based iterative localization. *IEEE Transactions on Parallel and Distributed Systems*, 21(5):631–640, 2010.

[122] K. Yu and Y. J. Guo. Statistical nlos identification based on aoa, toa, and signal strength. *IEEE Transactions on Vehicular Technology*, 58(1):274–286, 2009.

[123] H. M. Zhang. A theory of nonequilibrium traffic flow. *Transportation Research B*, 32(7):485–498, 1998.

[124] Y. Zhang. on theory of compressive sensing via ℓ_1-minimization: Simple derivations and extensions. *Technical Report TR08-11, Department of Computational and Applied Mathematics, Rice University, Houston, Texas*, 2008.

[125] F. Zhao and L. J. Guibas. *Wireless Sensor Networks: An Information Processing Approach*. The Morgan Kaufmann Series in Networking. ELSEVIER, 2004.