3-11-2011

Mitigating Congestion by Integrating Time Forecasting and Realtime Information Aggregation in Cellular Networks

Kai Chen
Florida International University, kchen001@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

MITIGATING CONGESTION BY INTEGRATING TIME FORECASTING AND

REALTIME INFORMATION AGGREGATION IN CELLULAR NETWORKS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Kai Chen

2011

To: Dean Amir Mirmiran
      College of Engineering and Computing

This dissertation, written by Kai Chen, and entitled Mitigating Congestion by Integrating Time Forecasting and Realtime Information Aggregation in Cellular Networks, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Kia Makki

_____
Stavros V. Georgakopoulos

_____
Deng Pan

_____
Yi-Min Zhu

_____
Niki Pissinou, Major Professor

Date of Defense: March 11, 2011

The dissertation of Kai Chen is approved.

_____
Dean Amir Mirmiran
College of Engineering and Computing

_____
Interim Dean Kevin O'Shea
University Graduate School

Florida International University, 2011

DEDICATION

I dedicate this dissertation to my family. Without their consistent encouragement and help, the completion of this dissertation would not have been possible.

ACKNOWLEDGMENTS

First of all, I want to show my gratitude to the Telecommunications and Information Technology Institute (it2) at FIU for providing me a pleasing and family oriented research environment. I am indebted to all faculty, researchers, and schoolmates of it2 who have encouraged, advised, helped, and supported me throughout the course of this work.

I would like to express my highest and sincere gratitude to my major advisor Prof. Niki Pissinou. She gave me a lot of constructive academic advice and helped me to mend my misguided approaches through my Ph.D. stage at FIU. She also made a lot of technical contributions to my dissertation work. Additionally, her attitude towards research and life sets up a role model for me and insures me to always be focused and never deviated from the objective of research. It would be a huge mistake if I fail to mention the name of Prof. Kia Makki, who gave me a hand when I was in desperation. Without him, I would not had a chance to start my research at the first place. I sincerely thank him for inspiring, supporting, advising and encouraging me towards my research goal. My special appreciation and thanks to all the members of my dissertation committee – Dr. Stavros V. Georgakopoulos, Dr. Deng Pan and Dr. Yi-Min Zhu, for their consistent support, patience and professional instruction.

I am extremely grateful to my family who backed me and supported me throughout the research. Without my wife's love and prayers, my father's guidance and my mother's encouragement, it would have been difficult for me to conduct this research.

Finally, I want to show my respect and appreciation to the Graduate School of FIU for providing the Dissertation Year Fellow which supported this research.

ABSTRACT OF THE DISSERTATION

MITIGATING CONGESTION BY INTEGRATING TIME FORECASTING AND

REALTIME INFORMATION AGGREGATION IN CELLULAR NETWORKS

by

Kai Chen

Florida International University, 2011

Miami, Florida

Professor Niki Pissinou, Major Professor

An iterative travel time forecasting scheme, named the Advanced Multilane Prediction based Real-time Fastest Path (AMPRFP) algorithm, is presented in this dissertation. This scheme is derived from the conventional kernel estimator based prediction model by the association of real-time nonlinear impacts that caused by neighboring arcs' traffic patterns with the historical traffic behaviors. The AMPRFP algorithm is evaluated by prediction of the travel time of congested arcs in the urban area of Jacksonville City. Experiment results illustrate that the proposed scheme is able to significantly reduce both the relative mean error (RME) and the root-mean-squared error (RMSE) of the predicted travel time. To obtain high quality real-time traffic information, which is essential to the performance of the AMPRFP algorithm, a data clean scheme enhanced empirical learning (DCSEEL) algorithm is also introduced. This novel method investigates the correlation between distance and direction in the geometrical map, which is not considered in existing fingerprint localization methods. Specifically, empirical learning methods are applied to minimize the error that exists in the estimated distance. A direction filter is developed to clean joints that have negative influence to the localization

accuracy. Synthetic experiments in urban, suburban and rural environments are designed to evaluate the performance of DCSEEL algorithm in determining the cellular probe's position. The results show that the cellular probe's localization accuracy can be notably improved by the DCSEEL algorithm. Additionally, a new fast correlation technique for overcoming the time efficiency problem of the existing correlation algorithm based floating car data (FCD) technique is developed. The matching process is transformed into a 1-dimensional (1-D) curve matching problem and the Fast Normalized Cross-Correlation (FNCC) algorithm is introduced to supersede the Pearson product Moment Correlation Co-efficient (PMCC) algorithm in order to achieve the real-time requirement of the FCD method. The fast correlation technique shows a significant improvement in reducing the computational cost without affecting the accuracy of the matching process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADAS   Advanced Driver Assistant

AID   Automatic Incident Detection

AMPRFP   Advanced Multilane Prediction based Real-time Fastest Path

ANN   Artificial Neural Network

ANOVA   Analysis of Variance

AOA   Angle of Arrival

AP   Access Point

ATIS   Advanced Traveler Information Systems

BFA   Believable Factor Algorithm

BTS   Base Station

CC   Cross Correlation

CDF   Cumulative Distribution Function

CDMA   Code-Division Multiple Access

CIR   Channel Impulse Response

DCSEEL   Data Clean Scheme Enhanced Empirical Learning

DF   Data Fusion

DFI   Direction Filter

DGPS   Differential GPS

DIKBA   Dijkstra's Approximate Buckets Algorithm

DRGS   Dynamic Route Guidance System

DSRC   Dedicated Short-range Communications

DT   Decision Tree

EARBALE	Environment-aware RSS-based Location Estimation
EFs	Environmental-dependent Factors
FCC	Federal Communication Commission
FCD	Floating Car Data
FDOT	Florida Department of Transportation
FFT	Fast Fourier Transform
FNCC	Fast Normalized Cross-Correlation
GCC	Generalized Cross-Correlation
GPS	Global Positioning System
GSM	Global System for Mobile communications
ITS	Intelligent Transportation Systems
KKT	Karush-Kuhn-Tucker
KNN	k Nearest Neighbor
LBS	Location Based Service
LLOP	Linear Lines of Position
LMU	Location Measurement Units
LOP	Line of Position
LS	Least Square
MAP	Multipath Angular power Profile
MF	Matched Filter
ML	Maximum Likelihood
MLP	Multi-Layer Perceptron
MPDP	Multipath Power Delay Profile

MS	Mobile Subscribers
MSE	Mean Square localization Error
MT	Mobile Terminal
NCC	Normalized Cross Correlation
NLOS	Non-Line-of-Sight
PDC	Personal Digital Cellular
PMCC	Pearson product Moment Correlation Co-efficient
PN	Pseudo Noise
PRFP	Prediction based Real-time Fastest Path algorithm
RBF	Radial Basis Function
RF	Radio Frequency
RISC	Rapid Incident Scene Clearance
RME	Relative Mean Errors
RMSE	Root Mean Square Error
RSS	Received Signal Strength
RSU	Roadside Unit
RTI	Real-time Traffic Information
Rwgh	Residual weighting Algorithm
SMS	Short Message Service
SoE	Sum of Exponential
SRM	Structural Risk Minimization
SS	Signal Strength
SVR	Support Vector Regression

TDMA	Time-Division Multiple Access
TDOA	Time Difference of Arrival
TFP	Travel time based Fastest Path
TOA	Time of Arrival
UMTS	Universal Mobile Telecommunications System
UWB	Ultra-Wideband
VANET	Vehicular Ad-Hoc Network
WSN	Wireless Sensor Network

CHAPTER 1

INTRODUCTION

1.1   Background

In a wireless network, the interconnection between each node is achieved without the use of cables. Generally, different electromagnetic waves (radio waves, microwaves, terahertz radiation, infrared radiation, etc.) are used as a carrier to transmit remote information for various applications. Wireless networks are widely utilized in people's daily life. For instance, people can communicate with each other through the use of a cellular phone and can connect to the internet in regions lacking of efficient telecom infrastructure.

As a special kind of wireless network, a cellular network consists of a large number of radio cells which are centered with a fixed base station to provide full duplex communication services. Multiple frequencies correspond to the radio base station are assigned to each single cell. In order to satisfy the sustained increase of the demand for network capacity and communication quality, the traditional analog cellular network was mitigated with some digital networks such as time-division multiple access (TDMA) and code-division multiple access (CDMA) in the United States, Global System for Mobile communications (GSM) in Europe, and Personal Digital Cellular (PDC) in Japan.

One primary characteristic of the cellular network is its ability to provide wide area coverage in order to allow a mobile terminal (MT) to move from cell to cell without loss of communication. This handover mechanism is achieved by the continually exchanging of signal pilots between the MT and surrounded base stations through the control channel.

Additionally, most environments where mobile units operate involve high-noise features. These noises may be caused by obstructions that block the transmission of radio signals, the frequent changes in atmospheric reflection, and the multipath propagation. The MT is power limited due to the rule of Federal Communication Commission (FCC) in the concern of the influence of human's health. A power control mechanism is utilized in the cellular network to reduce interference between other MTs and to conserve energy. Unfortunately, this mechanism also limits the observation of surrounded base stations which will degrade the performance of location estimation by reducing the number of base stations used in the positioning system.

Besides providing standard voice functions, a cellular network can also support many additional value added services, such as short message service (SMS) for text messaging, packet switching for accessing Internet, email, Internet gaming, and location-based service (LBS). LBS contains information collection, information sharing, and entertainment service that are supported by the mobile devices through the cellular network with its geographical position. For example, the position information of sampled MT placed in a vehicle is collected periodically in a short period in order to estimate the travel time of the arc where it is driving.

An Intelligent Transportation System (ITS) refers to the utilization of information and telecommunication technologies to support transport infrastructure. The fundamental interest of ITS comes from the demand for overcoming traffic congestion problems and providing synergy to new real-time control technologies. Based on the specific tasks it contributed, ITS can be categorized into advanced traveler information systems (ATIS), advanced driver assistant (ADAS), traffic demand estimation, traffic forecasting and

monitoring, network control, automatic incident detection (AID), and crash analysis and prevention.

In order to mitigate the congestion problem, the route guidance system is developed by operating a real-time process of guiding drivers along the route generated by a route planner. It integrates the general functions of ATIS, traffic forecasting and monitoring techniques to provide optimal route options for drivers to avoid involvement with congestions and to save traffic costs, such as time and gasoline. However, due to the complexity of the urban traffic network, the quality of traffic forecasting and monitoring services is volatile, and this will degrade the ability of route guidance system in determining proper routes.

Current research concentrates its efforts on developing various mathematical models to investigate the linear relationship between the predicted travel time and the historical traffic behavior data to improve the accuracy of the forecasting result. This can overcome the unintelligent traffic behavior of most drivers who is driving based on their empirical experience but not the real-time traffic conditions. Nevertheless, the influence of the fluctuation of traffic flow that was caused by congestions and incidents happened in adjacent arcs is not well investigated. This nonlinear affection may cause the observed traffic behavior to be far more different from the historical records which will introduce more prediction errors into present prediction models. Additionally, the cellular network becomes a good infrastructure for providing real-time traffic monitoring data due to its seamless coverage in metropolitan areas. Researchers are in the process of improving the accuracy of the estimated location result in outdoor cellular networks by designing complex radio propagation models or by developing new integrated components for

mobile devices. However, few of them realize that the learning theory based method, which already achieved good performance in location estimation in indoor environments, can be a good option in dealing with this problem. Utilization of these mapping approaches in the outdoor cellular network, especially the urban region, poses new challenges and problems. Therefore, both the unique characteristic of outdoor cellular channels and the application requirements have to be carefully considered.

This dissertation focuses on fusing the historical traffic behaviors and the real-time non-linear traffic patterns caused by neighboring arcs together to estimate the approximate travel time of an arc involved in congestion. We also concentrate on traffic information monitoring technologies in the cellular network and attempt to clean the collected position related information appropriately and improve the accuracy of the location estimation result.

1.2   Motivation

Traffic congestion has become a critical problem all over the world due to the fast development of cities and the persistent increase of motor vehicles. For instance, in the United States, the annual waste of gasoline is more than 10 billion gallons and the total economic loss caused by congestion is above 63 billion dollars. Additionally, the congestion will induce numerous environmental problems and human's health problems by extending the time period for drivers to finish their trips, which will cause the increase of the amount of automobile exhaust. Unfortunately, congestion cannot be simply solved by building more routes and widening the existing ones because of the limited living space, the lengthy construction period and the inestimable budget.

In the previous section, characteristics of travel time prediction based congestion mitigating systems were introduced for distinguishing them from other ITSs. The most significant feature of the travel time prediction based method is its ability to provide a benchmark that can represent current traffic conditions and balance the requirement of spatial and temporal efficiencies. However, unexpected incidents, irrational operations of vehicles, and the collaborations of conjoint routes weaken the significance of the historical records and degrade the accuracy of the predicted travel time. This data may mislead the system to provide improper guidance to drivers and loss their trusts.

For example, we know that an arc usually has a good traffic condition between 10:00 AM to 12:00 AM. However, one day there was an incident that happened and blocked part of the route. If at this time we try to predict the transit time of this arc based only on the historical traffic behavior without noticing the incident, the predicted time will be much shorter than the actual travel time to go through it. And the congestion avoidance system may select this route for drivers in order to increase their trip efficiency. However, this false decision cannot achieve this objective.

Although the real-time traffic patterns have big influences to the future travel time, obtaining the real-time traffic data in current transport infrastructure is hard to achieve. The traditional "in-situ" technology collects real-time traffic data through various kinds of sensors installed in the specific location along the highway. This technique can provide very accurate traffic data due to its direct access to the transport infrastructure. However, the cost of the deployment and maintenance of this system is extremely high that limited the coverage area only valid to some major freeways and highways. This means that only

a small part of routes are under monitored, and it is difficult to analyze the traffic conditions in adjacent arcs.

Different from "in-situ" technology, the Floating Car Data (FCD) technique gathers traffic status of probe vehicles travelling on each route to estimate its traffic speed and volume data. This method avoids the requirement of building complex detector systems which enables it to monitor even a tiny local arterial. The most popular applications of FCD are the GPS based approaches, which can locate the target probe within a 10m area. However, the finite distribution and the high power consumption of GPS-equipped devices constraint the FCD to achieve seamless monitoring of the urban traffic network. Therefore, as mentioned in the previous section, the cellular network supported FCD attracts people's attention. For instance, some companies and organizations like CELLINT, DELCAN, APPLIED GENERICS, AIRSAGE, and GLOBIS DATA, developed numerous test projects for justifying the performance of it. However, the accuracy of the estimated position of the cellular probe is constrained by the high-noise wireless environment where most mobile units operate. The received signal suffered from the influence of various errors and increase difficulties of the inverse operation (to derive coordinates from the received signal). This is a critical problem since we cannot specify the source of the estimated traffic information without the precise location information.

Network-based location estimation of mobile devices in the cellular network is becoming increasingly popular for providing LBS such as enhanced-911, field worker deployment, and real-time traffic monitoring services [1]. In traditional network-based location estimation solutions, signal propagation models that are based on the geometric and statistical approaches were developed and widely used to approximately estimate the

path loss of the received signal. However, building an accurate and proper signal propagation model is difficult and sometimes impossible. This is due to the channel noises which attenuate the field strength and spread the obstructed path of the transmitted waves. The source of those noises includes absorption, scattering, reflection, refraction, diffraction, multipath, and Non-Line-of-Sight (NLOS), which are defined as the Environmental-dependent Factors (denoted as EFs) [2, 3]. In previous research [2, 4-6], the machine learning theory based (denoted as learning-based) approaches are capable of avoiding the design of the complex propagation model and reducing localization errors. This is achieved by using the empirical data observed in the training phase where historical data in the same situation is collected before the estimation.

The motivation behind this research originates from challenges and problems declared in the above arguments involving the optimization of approaches of travel time forecasting, mobile unit location estimation, and cellular probe tracking.

1.3   Problem Statement

This dissertation originates from the observation that the conventional travel time forecasting approaches only focus on investigating the relationship between the future travel time and the historical traffic patterns (daily, weekly, and seasonal). These approaches ignore the collaborations exist in the transport architecture which will also affect the performance of future traffic behaviors. Additionally, the inefficient diversion strategy problem is not well considered by them. Therefore, the main aim of this dissertation is to design a congestion mitigating system that fuses both the historical

traffic behavior and the real-time non-linear traffic patterns of neighboring arcs together to overcome the limitation of present congestion mitigating systems.

The other objective of this research is to design and develop schemes that will reduce errors which exist in the received signal strength and the travel time using learning-based methods. On comparison with the conventional learning-based methods, these new schemes adapt the nonlinear expansion of the input vector and map this vector to a high-dimensional feature space. The Signal Strength (SS) profile of the received signal is utilized together in location estimation, particularly in urban areas where establishing a signal propagation model is complex and difficult. A learning-based method is used to predict and abate errors that exist in the SS data. Also, the developed scheme should contain a reliability feature required to be deployed in any real-time situation.

1.4   Significance and Contribution

Travel time as a parameter that does not belong to the nature of an arc has the ability to balance between multiple traffic flow properties, such as route length, traffic speed, and event issues. Using accurate travel time, especially the future time, we can provide a high quality weight parameter for congestion mitigating system to generate optimal routes for drivers to reach their destination quickly and smoothly. Travel time forecasting is more valuable in the congestion situation where calculating accurate travel time is difficult. This dissertation focuses on improving the performance of the fusing phase in travel time prediction according to the characteristics of transport infrastructure. It paves a way for the next generation congestion mitigating system that is able to adapt to the complexity and mutability of urban traffic networks.

8

The major contribution of this dissertation is the development of a novel congestion mitigating system for both the single lane and multi-lane urban environments. This system establishes a data fusing technique that consolidates interactions between the current arc and its surrounding infrastructures with the arc's historical traffic features to predict the prospective traffic conditions of this arc. This dissertation empirically illustrates that the congestion reason analysis scheme and the iterative travel time forecasting mechanism promote the accuracy and efficiency of the congestion mitigating system.

Furthermore, achieving high accuracy in the location estimation is significant for providing real-time traffic information to support the congestion mitigating system. This is difficult due to the complexity of the wireless environments. A large number of non-linear effects such as the NLOS propagation and the multipath propagation [3] exist there. Traditional solutions to the above open issues offer complex mathematical approaches to reduce location errors relying on physical parameters of the signal received by the base station. Using learning-based methods to solve the position estimation problem in the cellular network has been proposed in [2, 4-6]. This is a significant breakthrough since there is no need to build a specific and complex signal propagation model. Wu et al. [2] has identified several drawbacks such as the random and incomplete characteristics of a cellular signal, and the fluctuated wireless environments. They are critical to the successful implementation of machine learning algorithms in a realistic scenario. Without a proper solution to these drawbacks, it is difficult for the existing models to efficiently locate mobile objects. In this research, we attempt to design and develop new schemes to eliminate these drawbacks.

At the end of this dissertation, we extend the correlation based cellular probe speed estimation mechanism, that first specified by Chandrasekaran [7], by utilizing the Fast Normalized Cross-Correlation (FNCC) algorithm [8] to replace the Pearson product Moment Correlation Co-efficient (PMCC) algorithm [9] applied in the matching phase. This modification is critical in improving the efficiency of correlation based FCD and contributes to meet the real-time feature of the congestion mitigating system.

1.5 Methodology

This dissertation addresses the problem of accurately predicting the travel time of the congestion involved arc with the support of real-time traffic stream of adjacent areas collected from the cellular FCD system. The work demonstrates that the identification of the traffic status parameter through the traffic pattern classification model and the multi-lane congestion analysis model is necessary in determining a proper method to estimate the future travel time. Support Vector Regression (SVR) is utilized as the kernel based predictor to forecast the travel time of the arc that was not influenced by conjoint routes. The proposed iterative architecture of the congested route travel time estimation process combines the prediction result of SVR which is based on the current arc's historical traffic behaviors and the transit time of anterior arc that induce this congestion, to achieve a proximal estimation result.

In order to obtain detail traffic status data from the nearby transport infrastructure (highway, local arterial, etc.), a correlation based FCD approach, relying on cellular network, is introduced. Conventional learning-based approaches can achieve very accurate position estimation in a challenging environment with a high quality training

dataset. One of the problems of them is that the geometry feature (latitude, longitude, etc.) of position-related parameters is not well considered. Thus, we developed a data clean enhanced empirical learning algorithm to predict the more precise value of the estimated distance that relied on the SS data between the target point and each reference base station (BTS) before the position estimation phase. According to these distances, a group of joints that contains the location information of the target point are generated using the trilateration method [10]. An original filtering method is developed to clean the unqualified joints.

Besides mobile unit tracking, this dissertation also dedicates itself to improve the efficiency of the correlation based FCD method in extracting speed information. This is achieved by transforming the calculation of the coefficient points in the searching range into a whole-to-part matching problem. The fast normalized cross-correlation (FNCC) algorithm [8] is utilized to diminish redundancies that exist in the calculation of the correlation coefficient.

The mathematical proof for the new schemes will be established and their efficiencies will be evaluated by different simulations. Matlab 7.1 and Visual Studio.Net 2008 are going to be used for synthetic simulation for this research work. The accuracy of predicted travel time, the new congestion mitigating system and localization schemes' efficiencies, robustness and their abilities to survive in imperfect environments will be demonstrated using different simulation results. The main parameter that needs to be measured and analyzed is the congestion ratio of the urban network, the standard deviation of errors of the predicted SS data, and the average time consumption of the matching phase. Other parameters used in evaluating the performance of our scheme

11

include: the mean absolute error (MAE), the relative mean errors (RME), the median localization error (MLE), the root mean square error (RMSE), the cumulative distribution function (CDF), and the histogram of the localization error distribution.

1.6   Organization of the Dissertation

The organization of this dissertation is defined as follows: In Chapter 2, an exhaustive literature survey about the acquisition of location related parameters and position estimation is done. Chapter 3 presents the cellular network supported congestion mitigating system with proper justifications to the claims made. Chapter 4 focuses on the novel data clean scheme enhanced empirical learning algorithm (DCSEEL) with appropriate evaluations of the improvement in the location estimation result. In Chapter 5, a theoretical procedure using FNCC to improve the computational efficiency of the one dimensional signal profile matching phase is presented. In Chapter 6, conclusions of the dissertation and summarization of the simulation results of this work are presented. Future works of this research are further discussed in this chapter.

CHAPTER 2

RELATED WORK

2.1   Introduction

A cellular network [11] is a radio network composed of hexagonal multiple shape radio cells in order to provide desired services to telephone customers. As a fundamental of Global System for Mobile communications (GSM) and Universal Mobile Telecommunications System (UMTS), cellular networks utilize various technological advancements in volatile wireless environments to provide services of mobility management, call set up, registration, and handover. In the contemporary world, the coverage of cellular networks across more than 212 countries and more than 3 billion people are benefited from their services all over the world.

One important service supported by the cellular network is the location-based services (LBS) which enables personalized services to mobile subscribers relying on their current position [4]. For example, a mobile terminal may be designed to detect the position of a subscriber, and is able to send advertisements of nearby shops to this customer. A critical issue in cellular network is the way the location related parameters are collected, analyzed and utilized to estimate an accurate position of a mobile terminal. Specifically, in the case of parameter analysis, the pivotal step is to cleanse nonlinear residuals that exist in received radio frequency (RF) signals. In this respect, cellular networks based location estimation studies as a discipline which involves extraction, estimation and rectification of spatial coordinates to overcome constraints of the harsh cellular environment. To sum up, location estimation in cellular networks can be defined

as the manner to calculate the concrete coordinates of a mobile terminal by calibrating and deriving erroneous signals.

Location estimation is an emerging application where noisy position-related parameters are measured from multiple channels to derive the location of the mobile unit. The first application of location estimation of the mobile terminal can be retrospect to World War II [12]. It was designed for accurately and rapidly tracing the military personnel in emergent battlefields. Due to the dramatic increasing of the demand of location awareness, emergency call location, and ambient intelligence services, accurate location estimation has become imperative to support new generation of LBS.

The primary challenging aspect of location estimation in cellular networks is to impute the channel noises which attenuate the field strength and spread the obstructed path of the transmitted waves. The source of those noises includes absorption, scattering, reflection, refraction, diffraction, multipath, and Non-Line-of-Sight (NLOS), which are defined as the Environmental-dependent Factors (denoted as EFs) [2, 3]. In the ideal situation, after the execution of location estimation, the geometrical coordinates of the target should be solely identified, released from any discrepancies, and made adaptable for supporting any applications of the LBS. However, there is no perfect location estimation mechanism for cellular networks so far due to the time-vary nature of wireless environments [5], the lacking of complete environmental factors [2], and the unaffordable cost of integrating new components into the development of mobile terminal [10].

2.1.1   Terminology

Generally, location estimation can be operated with two steps: position related

parameters estimation and position estimation. In each step, many operations are encompassed respectively. Some examples of these operations are shown below:

- Signal Parameter Extraction

- Parameter Discarding

- Parameter Replenishing

- Dataset Training

- Propagation Model Determination

- Regress Scheme Learning

- Location Approximation

In this study, the term "location estimation" will be used, which includes the basic operations of position related parameter estimation, propagation model determination and regress scheme learning. Furthermore, terms such as data, signals, and parameters are also used in this dissertation, to refer the values used for locating the position of the mobile terminal.

2.1.2   Shortcomings of the Outdoor Cellular Network

Generally, the cellular network is designed to provide services in complex environments especially in the outdoor region which encompasses urban areas, suburban areas, and rural areas. The signal may be transmitted from far away base stations and experiences unpredictable fluctuations due to the frequent change of the characteristics of RF channels. For instance, a signal may be bounced off mountains, buildings or other obstructions during its transmission, which poses multipath problems for representing precise spatial correlations. This is caused by the inability of the receiver to distinguish

the reflected signal with the straightly traveled one, and the accuracy of the estimated position is confined. Additionally, even in the ideal condition where all signals can penetrate obstacles without reflection, they may still not be completely observed by receivers. This is due to the non-line-of-sight (NLOS) propagation problem, that signals can be absorbed or garbled by those obstacles.

The cellular network also faces constraints of the system. For example, not all the base stations are installed with mutual synchronization equipments, which means that their signals do not enclose with timing information. Thus, the accuracy of location estimation using these data is degraded remarkably. Otherwise, since the primary demand of the cellular network is providing continual telephone services to mobile subscribers, only a small subset of signals propagated from base stations are large enough to be readable on a mobile terminal. Additionally, the bandwidth limitation in some kinds of cellular networks (GSM) also constrains the quality of received signals.

2.1.3 Outline of the Chapter

The purpose of this chapter is to provide a survey on location estimation mechanisms and architectures for the cellular network, and to discuss the relationship between them. The background of this dissertation is presented in the following structure: Section 2.1 presents the terminology, fundamentals and significance of location estimation; Section 2.2 discusses different methodologies of location estimation presented in the literature in the cellular network especially in the outdoor environment; In Section 2.3, a summary of this chapter is provided.

2.2   Location Estimation

Location estimation in the cellular network is designed specifically for conquering the drawbacks and shortcomings mentioned in Section 2.1.2. Along with these drawbacks and shortcomings, the primary challenge of location estimation in the cellular network is that the estimation process has to derive accurate position information from a stream of retrievable data which is incomplete and deviated.

Generally, location estimation can be divided into two categories: handset-based location estimation and network-based location estimation. For handset-based location methods, the core idea is to locate the handset equipment with the help of an integrated GPS unit [13]. Despite the high accuracy of the localization result, high cost and long periods of deploying new handset equipments in the whole cellular networks lead the handset-based method cannot be a general location estimation solution for supporting all kinds of applications of LBS. For instance, in terms of engineering view, to upgrade millions of mobile phones which had already been in operation is not efficient and economical. Additionally, the poor performance of GPS in indoors or urban canyon environments where tall buildings are densely distributed also constrains the implementation of handset-based location estimation methods.

A large number of network-based location estimation approaches have been developed in literature to calibrate the location of mobile units using the harsh RF signals in the cellular network. The common two-step architecture is shown in Figure 2.1.

2.2.1   Position Related Parameters Estimation

In order to extract spatial data from a received signal, the first step is to identify and

17

estimate position related parameters. Generally, signal parameters that can be employed in location estimation encompass power, timing, and direction of a received signal.
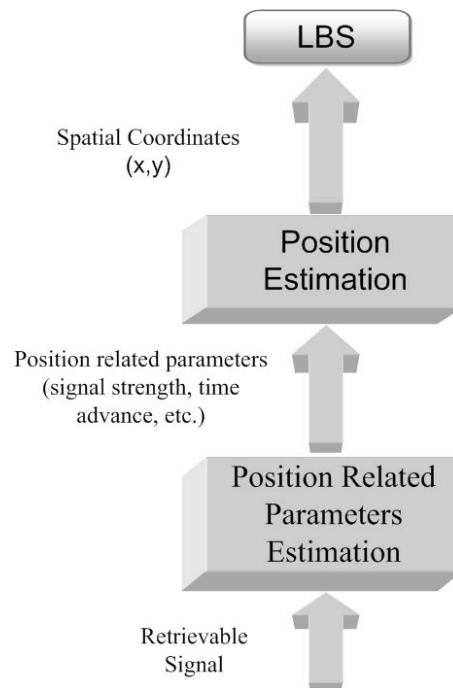
Figure 2.1: Two-step Architecture of Network-based Location Estimation [10]

A. Signal Strength (SS)

Signal strength can be defined as the power or energy of a signal traveling between the mobile terminal and base stations. As a signal parameter, it contains information that relates to the distance or range between the cellular subscriber and the served base station. When combining this parameter with a properly designed path-loss and shadowing model, the distance d can be approximated. Thus, in the ideal situation, the location of MT can be concluded on the circle in which the center is the serving base station and the radius is the calculated distance for two-dimensional positioning (as shown in Fig. 2.2).

The main advantage of SS is that the signal strength, or the signal attenuation parameter, is the only common information available among various kinds of cellular

network [4]. The economic and compatible features of SS [14-16] also draw great attention of researchers due to SS's ability to avoid installing expensive components in each side.
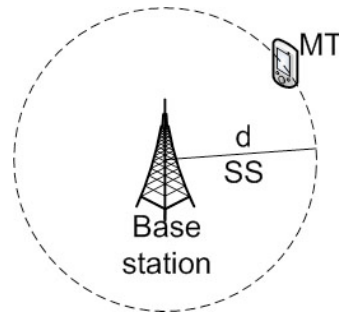
Figure 2.2: One base station measures the SS and determines the distance d between itself and the MT (uncertainty is defined as a circle)

However, in practice, an error-free environment is not achievable, especially in the outdoor urban region. This means that the signal exchanged within cellular communication networks has to experience fast (multipath) fading, shadowing and path-loss, which may lead to a complex signal shape. These influences become rapidly serious along with the increase of the distance. Another drawback of SS is its random and incomplete characteristics, which raise a great challenge to the location estimation [2]. These are caused by the original settings of the cellular network: only a small subset of RF signals can be recorded on a mobile handset and their strength need to exceed the threshold of the smallest handset reading.

B. Angle of Arrival (AOA)

In addition to developing new method to overcome the shortcomings of SS, a new parameter was proposed to serve location estimation. The AOA parameter is derived from the signals received by different antenna elements which are installed in the base

station or the handset [10]. With this parameter, a unique measurement will create a straight line to connect the base station to the MT. Thus, if an antenna array is installed and with known array geometry, the time differences of the same arrival signal can determine the angle between the base station and the MT [17]. For instance, in a narrowband communication environment, the phase shift is utilized to present the difference of the arrived time. Then, the AOA can be determined by using this shift data combined with the geometrical feature of each antenna element [18].

The AOA scheme can provide enough location information to rely on a minimum of only two base stations. However, the requirement of installing antenna array is difficult to be implemented and is not economical. Additionally, the accuracy of AOA measurements is highly related to the cell range. A tiny error in the estimated angle can be amplified to a large position deviation in urban environments where MT is far away from base stations.

C. Time of Arrival (TOA) / Time Difference of Arrival (TDOA)

The TOA parameter is defined as the flight time of a signal transmitting between MT and a base station. The estimation of this parameter is similar to the estimation of the SS parameter which is focusing on collecting the distance related information. This is due to the assumption that most RF signals are traveling at the speed of light (3×10^8m/s). Thus, in the absence of errors situation, the position of the MT will be laid on the circular locus around base stations (similar to Fig.2.2). In order to improve the quality of TOA, a synchronized clock is necessary to provide a unified time tick for an exchanged signal. For a two-way communication situation, exchanging time information by certain protocols is also an efficient way to maintain the synchronization feature [19].

Turin et. al. [20] firstly utilized the phase estimation method to obtain a TOA

parameter by employing a phase detector in his simulation. However, phase estimation method requires synchronization from at least three base stations, which sometimes cannot fulfill in practice. In 1992, H. Hashemi [21] proposed a pulse transmission method for wideband cellular network TOA parameter estimation. In this method, a correlator-based technique is applied to correlate the received signal with a local template for diverse delay corresponding, which can be calibrated based on the correlation peak. Meanwhile, a similar approach using the matched filter (MF) receiver to estimate TOA was also mentioned in Hashemi's research [21]. The filter engaged in this method matched with the signal transmitted between MT and a base station and executed estimation when the output of MF reached the maximum value. Both of these schemes can achieve optimal results in the maximum likelihood (ML) sense [22]. However, the performance of them degrades sharply in the multipath environment where the received signal is used as the template signal or the MF impulse response. Apart from phase estimation and pulse transmission techniques, the TOA can also be concluded through the spread spectrum technique [23, 24] in literature.

The main shortcoming of the estimation of the TOA parameter is the imperative of an installed synchronization clock which is expensive and hard to be achieved. Additionally, the accuracy of TOA data is influenced by the multipath problem and the NLOS noise, which are commonly observed in an urban environment.

In order to overcome the lacking of synchronization equipments between base stations and MT, an advanced method of signal travel time estimation, called time difference of arrival (TDOA) was developed, where synchronization is unavailable among each reference base station [18]. TDOA estimation calculates the time difference

of two arrival signals received by two base stations from the MT. After subtracting these two different values, the synchronization error (also called timing offset [10]) between MT and base stations can be eliminated, and then used to locate the position of the MT on a hyperbola (on which foci are at the two reference base stations and shown in Fig. 2.3). This is due to the assumption that all the synchronization error between the unsynchronized pair of MT and each base station are the same.
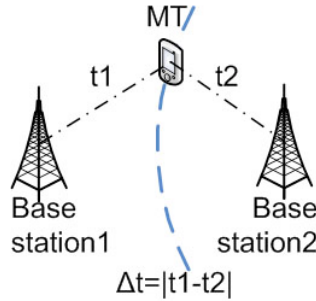
Figure 2.3: TDOA hyperbola based on the time difference between MT and two base stations

Another method of TDOA parameter estimation developed by Caffery and Stüber [25], as an extension of their previous work, is focusing on calculating time difference value of two reference signals with the help of cross-correlations. In this method, the largest cross-correlation value is observed and applied in the calculation of the corresponded travel time delay. The formula of cross-correlation-based TDOA estimation is

$$R_{t_1 t_2}(\tau) = (t_1 \star t_2)(t) = \frac{1}{T} \int_0^T t_1(t) t_2(t+\tau) d\tau \qquad (2\text{-}1)$$

$$\tau_{TDOA} = \arg\max_{\tau} R_{t_1 t_2}(\tau) \qquad (2\text{-}2)$$

where $t_i(t)$, for i=1, 2, … represents the travel time of signal between MT and the base station i, and T is selected as the observation time interval. Unfortunately, this

correlation-based method can only achieve good performances in single path channels where white noise is the primary interference. This performance will be degraded when signals transmitting through multipath channels with colored noise exist.

Aatique, in his master thesis [26], proposed an advanced method that utilizes generalized cross-correlation (GCC) techniques [27] to cross-correlate the filtered signals. This method is able to mitigate colored noise in multipath channels by shaping the cross-power spectral density that corresponds to the filtered versions of the signal. It can provide more rigorous TDOA results than the conventional peak detection method. However, the performance of this method is constrained seriously when deriving information from microphone signals.

D. Hybrid Schemes

Since the single position related parameter has various constraints in providing high quality localization services, hybrid methods which can provide extra position information of the MT, had drawn continuous attention from different scientific communities. The hybrid TOA/AOA parameter is used in location estimation by Aatique et. al. [28] in order to provide accurate position when there is only a single base station available to serve the MT. Different from TOA only and AOA only methods, with a hybrid identification of the LOS base station makes the new hybrid method is possible to eliminate the NLOS influences. The drawback of this hybrid scheme is that the improvement of accuracy is not prominent compared to others. However, its ability to provide location estimation in a lacking base station coverage situation without modifying the handset still makes it a good option in rural regions.

In order to further increase accuracy of location estimation when there is no

constraint in the number of serving base stations, a hybrid TDOA/AOA scheme which encompasses the advantages of both TDOA and AOA parameters was proposed by Li et. al. [29]. One advantage of TODA/AOA method is its ability to obtain the parameter of both AOA and TDOA at the same time. The AOA information can locate MT with only two base stations as the minimum, while the TDOA data can achieve a high accuracy when three or more base stations are available for location estimation. The AOA information can be used to compensate the TDOA parameter in order to solve the ambiguity problem of the TDOA approach. The other advantage is that the TDOA information can also help the AOA only method to avoid its drawbacks. In [17], we learned that the AOA information cannot be measured at the MT due to the uniformly distributed characteristic of AOA and the requirement of installing complex antenna array. Additionally, the near-far effect, which can degrade the quality of the neighboring base station's received signals due to the power control mechanism of the reverse-link pilot signal sent from the MT, resulting in low accuracy of AOA measurement. Furthermore, to gather the AOA information of the reverse-link pilot signal from the MT, antennas installed in each base station should be able to keep tracking the movement of a MT. Generally, this operation will introduce huge burden on the antennas. The TDOA measurements can help the AOA scheme to reduce the impact of near–far effect by applying the soft handoffs of the signaling and pseudo noise (PN) code tracking at the MT receiver, where the forward pilot signals are not constraint by power. Fourth, the TDOA scheme can release the synchronization requirement between MT and base stations, in contrast to the TOA approach. However, in order to achieve high localization accuracy, a long period TDOA/AOA measurement is necessary because of the

continuous transmission property of the forward and reverse pilot signals. Furthermore, special attention needs to be paid to ensure that error from either one method will not influence the overall performance of location estimation.

Along with the development of TOA/AOA and TDOA/AOA schemes, the other important hybrid parameter, TOA/TDOA, was proposed by Reza in his master thesis [30]. In his research, the raw measurements of TOA and TDOA are collected respectively. Since they have different error biases and variances, the weight of TOA and TDOA measurements that rely on their quality make it possible to improve the general performance of the estimated location.

Catovic and Sahinoglu [31] proposed a novel scheme using hybrid TOA/RSS or TDOA/RSS scheme to extend TOA only, TDOA only or RSS only method. They concluded that using the estimation result of RSS in conjunction with TOA or TDOA parameter can enhance the location estimation in short-range communications. Two enhancements can be achieved by TOA/RSS or TDOA/RSS schemes, including improvement of the overall location estimation accuracy and the distinct promotion of the nearby reference devices' performance. However, these methods can only improve the accuracy in short-range communication networks such as wireless sensor networks (WSNs), Wi-Fi, Bluetooth, and Ultra-wideband (UWB). Along with the increase of coverage area, the performance of TOA/RSS and TDOA/RSS become close to TOA and TDOA, gradually.

E. Others

Apart from the conventional position related parameters mentioned above, few other

researchers concentrated their efforts on analyzing the exchanged RF signal and finding new parameters for specific applications.

Multipath power delay profile (MPDP) [32] was proposed to be utilized in cellular networks location estimation based on the assumption that the characteristic of all multipath components' excess delay is resolvable. Nevertheless, MPDP can achieve good performance only in an indoor adverse environment where multipath and NLOS problems are serious. And a high time resolution is necessary to support the MPDP approach. Similar to the MPDP scheme, a multipath angular power profile (MAP) parameter was mentioned by the same author [32]. It focuses on estimating the multipath power profile at each node using antenna arrays.

In order to accomplish rough location estimation for some specific applications within the UWB channel, a channel impulse response (CIR) parameter, which exploits the nature of UWB channels, was nominated by Althaus, Troesch, and Wittneben [33]. In this scheme, the CIR of a transmitter and receiver pair (TX/RX) is unique. The CIR received by a RX works like a signature of any TX. If two CIRs are similar, it means the two TXs are close to each other. This feature can be used to decide the geographical region of the target node.

The primary advantage of these new developed parameters is their ability to provide more position related information in location estimation. In contrast, one drawback of these schemes is the requirement of building a database to store preceding parameters in order to extract location information which will increase complexities of the position related parameter estimation phase. The other drawback is that most of these schemes are only efficient for selected applications.

2.2.2   Position Estimation

After the first step, which collects position related parameters, the next step is to design an appropriate method to extract position information from them, to reduce variants, and to calculate the position of the MT. Generally, this phase can be categorized into two different types based on the availability of database technologies.

A. Geometrical and Statistical Techniques

Geometrical Techniques

In a nutshell, the geometrical technique is defined as the technique that uses a deterministic approach to do position estimation relying on certain geometric relationships gathered from the first step.
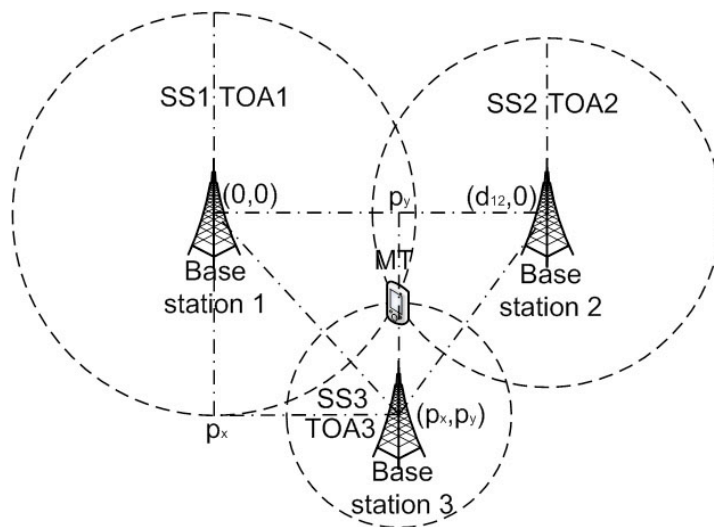
Figure 2.4: Trilateration geometrical technique for SS or TOA

The trilateration method [34] is the first geometrical technique derived from a mathematical solution in order to solve the three-dimensional trilateration problem, which was proposed by a Dutch mathematician and astronomer Willebrord Snellius. This method is able to determine the intersections (the final position of the MT) of three or

more globe surfaces with given centers (the known position of each base station) and radii (the distance decided by the SS or the TOA parameter which was extracted in the first phase). The operation of the trilateration method is shown in Fig. 2.4.

For the AOA parameter, the triangulation approach, which processes location estimation based on measured angles from reference nodes to the target node, is applied. In this method, the distance information is excluded in position estimation, but the angle data from two fixed reference nodes is enough to decide the location of target node, as shown in Fig 2.5.
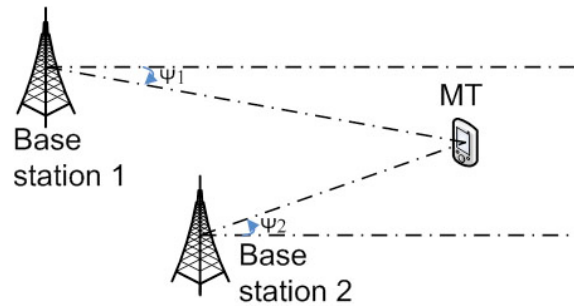
Figure 2.5: Triangulation geometrical technique for AOA

In order to determine the intersection of two hyperbola where the target unit foci on, a multilateration approach was developed to do location estimation relying on the TDOA parameter. The mutlilateration method requires data exchange between MT and at least three base stations. Generally, the multilateration method can achieve far more accurate location results, in contrast to sparse methods (trilateration, triangulation). The location estimation procedure is shown in Fig. 2.6.

However, due to the existence of variances and Environmental-dependent Factors (EFs) such as multipath and Non-Line-of-Sight noises in imperfect wireless channels in the cellular network, location related parameters collected from the first step are not

always error-free. Thus, the direct position estimation depending on these parameters may not be able to determine a unique intersection. Therefore, using geometrical techniques in non-ideal wireless environments can only provide an approximate region of where the MT may be located. The other drawback of geometrical techniques is the lacking of efficient data fusion mechanism which makes the effective utilization of multiple parameters estimation hard to achieve.
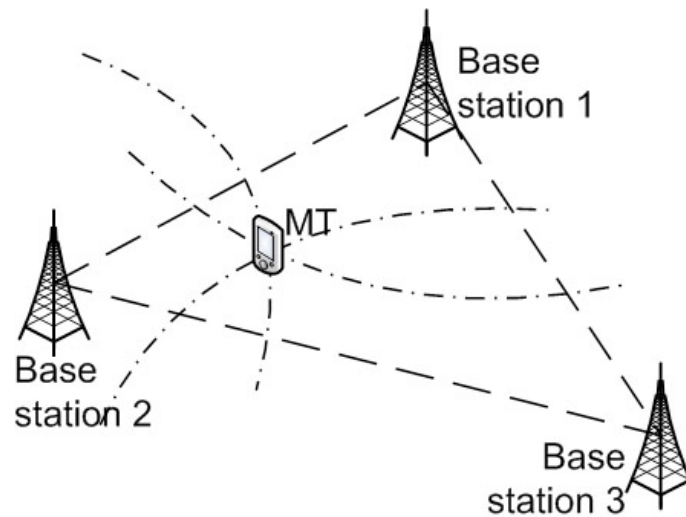
Figure 2.6: Multilateration geometrical technique for TDOA

Statistical Techniques

Similar to geometrical techniques, the statistical approach is also focusing on calculating the position of the MT. However, it acquires the most likely position of the MT from position related parameters with the presence of a theoretical framework. The utilization of the theoretical framework is the essential difference between geometrical methods. The fundamental model of the statistical approach is named as generic framework in (2-3, 2-4).

In (2-3), p_i is denoted as the position related parameter and function $P(x_i, y_i)$ is the

relationship model for the true spatial coordinates of the target node and the parameter. η_i is the noise parameter for current estimation. In (2-4), there is the definition of relationship model in (2-3) for different position related parameters, respectively. In an error-free situation where $\eta_i=0$, the generic framework can perfectly calculate the true coordinates of the MT.

$$p_i = P(x_i, y_i) + \eta_i \qquad (2\text{-}3)$$

$$P(x_i, y_i) = \begin{cases} \sqrt{(x - x_i)^2 + (y - y_i)^2} & \text{RSS/TOA} \\ \tan^{-1}\left(\frac{y - y_i}{x - x_i}\right) & \text{AOA} \\ \sqrt{(x - x_i)^2 + (y - y_i)^2} - \sqrt{(x - x_j)^2 + (y - y_j)^2} & \text{TDOA} \end{cases}$$

$$(2\text{-}4)$$

Unfortunately, the noise-free environment does not exist in the real world. Thus, researchers contributed their efforts to accurately determine the noise term η in order to extend the generic framework. In [35-37], the serious influence of NLOS error to the accuracy of location estimation was evaluated and defined as the "killer issue". In order to achieve high accurate location estimation result, it is necessary to correct the NLOS error a priori to location estimation.

NLOS Mitigation Methods

The first solution for mitigating NLOS errors was proposed by Wylie and Holtzman [38]. The first step designed in this method is a simple hypothesis testing for NLOS errors identification. After confirming the existence of NLOS ranging errors, a correction scheme is used with the assumption that there is prior knowledge of the standard measurement noise of the wireless environment. Wylie and Holtzman's work verified the possibility of identifying and mitigating NLOS errors utilizing some simple algorithms.

However, this correction solution is so naive that it can only achieve limited improvement in NLOS ranging errors mitigation. The other drawback of this method is its requirement of range measurements from a user as a time series. In order to work efficiently, the standard deviation of NLOS range measurements needs to be greater than the standard deviation of LOS measurements.

Another method proposed by Xiong et. al. [39] concentrated their efforts on reducing NLOS errors with the AOA parameter. The selective model identified and reduced the NLOS error relying on the coordination of many base stations. It first estimates the coordinates of MT using the LOS propagation model. Secondly, the error of the AOA parameter at each base station, its absolute value, and the root mean square of all errors are calculated. Then, every base station with large error value will be excluded from the sample list and the location of MT is recalculated using the maximum likelihood (ML) model [40-42]. The drawback of this method is that the accuracy of location estimation still cannot satisfy the requirement of FCC. Additionally, it fails to mitigate NLOS errors when more than one NLOS base station exists in the cellular network.

In order to overcome the limitation of multiple NLOS base stations situation, Chen et. al. [43] proposed a Residual Weighting Algorithm (Rwgh) that used deletion diagnostics to compute and rank the influences of excluding NLOS base stations from the entire set. The main contribution of Chen's work is its ability to mitigate the NLOS error when lacking NLOS measurement information. The primary drawback of Rwgh is the constraint that its performance will be fully degraded when all the propagation between MT and each base station is NLOS.

New approaches, using the robust estimation theory such as the pseudo-median

estimator [44] and range measurement model [45], are able to smooth the influence of NLOS errors with the integration of an estimator which is insensitive to small number of outliers. In these approaches, the location of MT is assigned with a unique maximum value which can be determined by the given path loss measurements between MT and three or more base stations, and the likelihood function [44]. The estimated location that has the largest likelihood value is determined as the final location of MT.

However, these robust methods face some constraints in a practical cellular system. First is that the communication channel between MT and most base stations have NLOS errors involved, especially when MT is far away from base station. The other drawback is that the number of base stations is often limited, and it is difficult to satisfy the requirement of the above methods. Thus, the quality of the likelihood ratio may be degraded.

During the same period, Caffery et. al. [46] proposed a Linear Lines of Position (LLOP) algorithm to divide the nonlinear intersection problem into several linear sub problems in order to reduce the complexity of calculation. This novel method focuses on solving the NLOS problem which exists in the generic TOA based geometrical techniques. However, the performance of this algorithm depends on the relative location of MT and base stations. Furthermore, to achieve reliable results, at least four base stations are necessary to provide straight lines of position (LOP) data. This is difficult to implement in the cellular network.

The Kalman filtering technique [47, 48] is first utilized in mitigating NLOS errors that exist in location estimation by Thomas, Cruickshank and Laurenson [49]. In their research, a robust architecture based on the biased Kalman pre-filtering of the TOA

32

parameter is designed as a crucial stage. It provides a real time TOA variance estimation which is able to extrapolate TOA value when the LOS path is lost. The first weak point of this Kalman filtering method was the use of the standard deviation of a Kalman filter's estimation as the identification of the LOS/NLOS scenarios. Additionally, the simple combination of the using of this standard deviation and the increasing of the size of the covariance matrix of measurement noise vector is not efficient in mitigating NLOS error. These are due to the dynamic change of propagation channels in the real world where most mobile subscribers keep moving and their movements are unpredictable. The uncertain change of the velocity and the acceleration, as well as the correlation of measurement data, makes it quite possible to generate mismodeling errors to the Kalman filter estimation and degrade its reliability. Thus, the true variance in LOS and NLOS scenarios may not be able to be calculated through the Kalman filter.

LOS range estimation used in NLOS environments is a novel method to reduce the NLOS error existing in location related parameters [50, 51]. A constrained nonlinear optimization scheme is the key in this method, and it can achieve good performance when only three base stations are available for range measurements. The geometry of the cell layout is used to analyze the bounds on the NLOS error and the relationship between the true ranges. In LOS range estimation, there is no requirement of the distinction between NLOS and LOS base stations. This means that at any time, the propagation between serving base station and MT can be LOS or NLOS, and the cumulative error in the range measurements of the other base stations is non-negative. Thus, the NLOS error can be eliminated by computing normalized scale factors which adjust the NLOS-corrupted range measurements close to their LOS values. Another advantage of

33

this method is that it does not require any modification to the subscriber equipment. However, the implementation of this method in practical systems is difficult due to the inefficiency of the nonlinear objective function. Furthermore, the improvement of this method in NLOS errors elimination cannot reach the FCC's requirement.

Channel-scattering model based statistical algorithms have been designed to provide location estimation service in NLOS environments. These models can be divided into the ring/disk of scatterers [52], the Gaussian scattering models [53], and the dominant scatterers [3] based on different scatterers used in the true range measurements estimation. These algorithms utilize the TOA parameter or TOA/AOA parameters and their distribution function with the channel-scattering model to calculate the true measured range. The primary advantage of the channel-scattering mode based location estimation approaches are that they only need three base stations in the system for providing location information. The channel-scattering mode also can be applied for the case of a stationary user. However, the applicable areas of this method are limited. Thus, it is necessary to characterize the scattering environment before applying the algorithm in a particular area.

The believable factor algorithm (BFA) based method was proposed by He et. al. [54] in order to mitigate the influence of NLOS. In this method, the geometrical feature of cell layout was combined with the path loss model, the received signal strength, and the time of arrival (TOA) range measurements. It only requires TOA data from 3 reference base stations and it does not care about the propagation of each base station, including NLOS error or not. The believable factor is determined by the combination of the received signal strength (RSS) and a specific path loss model. It decides the reliability of the range measurement by TOA.

The BFA approach proposed by He et. al. did not consider the situation when there is no intersection between each TOA circle. Thus, in 2008, an extension method which integrates BFA and Kalman filter techniques at the same time to further increase the quality of estimated location and to reduce the influence of NLOS error is proposed by M. Y. Shi et. al. [55]. In this method, after identifying the existing of NLOS propagation, the biased Kalman filter model is utilized first to smooth the NLOS error. Then, the believable factor based objective function is adopted in order to further optimize the NLOS mitigation and MT localization. In this method, both advantages of BFA and the Kalman filter model can be inherited. However, the drawback of this method which is similar to all the typical believable factor method is the requirement of the quality of the selected path loss model. The highly accurate location estimation may not be attained when the selected path loss model cannot well approximate the real propagation conditions.

B. Mapping techniques (Fingerprint techniques)

The core feature of the mapping technique that differentiates it from non-parametric statistical methods is the utilization of the training data set which consists of previous sample records to estimate the position of the target node. Those samples contain the estimated signal parameters and the known spatial coordinates. Generally, the training phase is offline and operated before the real-time localization. A regression scheme is determined relying on the training data set and used to calculate the position of a given node.

The first fingerprint technique, RADAR, for wireless LAN (WLAN) was proposed by Bahl et. al. [56] in order to locate and track users inside a building by adopting the

k-nearest-neighbor (KNN) model. In this method, the MT's location is determined by comparing multiple data sets of previous training data with the observed signal strength and selecting the best matched one. The primary contribution of this method is providing a generic method which combines empirical techniques with the signal propagation model to estimate the position of a target node. However, this method lacks the ability to handle noisy wireless channels, which is common in the real scenario. Additionally, it requires large search space in order to achieve good performance.

In order to improve the ability to tackle the noisy wireless channel, Youssef et. al. [57] proposed a joint clustering technique combined with a probabilistic method for in-building location estimation. The main contribution of Youssef's job is that he defined a way to describe the tradeoffs between the accuracy of location estimation and the requirements of computation. In order to implement this tradeoff, two techniques, Joint Clustering and Incremental Triangulation, are selected and utilized together. However, this method still cannot reduce the computational burden of the location determination algorithm at an application level.

In order to further improve previous mapping techniques, Battiti et. al. [58] proposed a multi-layer perceptron (MLP) architecture based neural network model. This is an efficient solution to estimate the position of the target node and to reduce distance error extracted from location related parameters by applying neural network models and automated learning techniques. The relationship between the observed signal strength parameter and the spatial coordinates of the mobile unit is analyzed by the neural network model based on a big number of training parameters and the nonlinear transformation of each unit. One advantage of this approach is that the detailed knowledge of each access

point (AP) is not required. However, the performance of neural network fluctuations depends on the structure of the network and the complexity of training samples. Additionally, deciding the number of hidden layer nodes is also difficult in the real world scenario.

A nonparametric kernel-based estimator for location estimation was proposed by McGuire et. al. [59]. The propagation delay is selected as the parameter which is measured and analyzed in this project. The environment of this approach is not limited to the indoor environment but a microcell situation with both LOS and NLOS radio propagation. The main contribution of this approach is the definition of the mean square error's lower bound of the estimated location result which the kernel-based estimator location estimation method can achieve. Compared to the previous parametric maximum likelihood estimator, this new estimator can mitigate NLOS error better. The influence of NLOS noise can be degraded by the robust kernel functions.

Driving from the kernel-based estimator in McGuire's work, Gezici et. al. [60], presented a Support Vector Regression (SVR) [61, 62] based algorithm that estimates the position of a mobile unit relying on collected training data from a priori locations. Compared with conventional empirical methods, SVR utilizes structural risk minimization principle to overcome the limitation of the empirical risk minimization principle by minimizing upper bound (the highest limited value) of the expected risk. As an advanced regression algorithm, SVR maps the input data series into high feature space in order to learn the nonlinear relationship between input and output. The primary advantage of the SVR based method is that it achieves very accurate estimation result when NLOS propagation is serious.

However, the main drawback of the machine learning theory based method is its assumption of high quality and time relevant training dataset, which is too difficult to be achieved in cellular network, especially the outdoor environment. In order to achieve a very accurate location result, the main underlying assumption is that the training set needs to remain relevant during the whole evaluation period. It means that the environment should be sufficiently stable so that the measurements are associated with similar error statistics.

A novel method which treats the location estimation problem as a nonlinear classification problem was proposed by Sun et. al. [63] in order to overcome the influence of NLOS propagation. This method is an extension of the conventional least square (LS) algorithm with advantages such as high-accuracy, real-time estimation, and fault-tolerance. Additionally, it can estimate the accurate position of MT without requiring ad-hoc infrastructure in addition to the cellular networks. The detailed knowledge about the locations of each base station and the environment characteristics has no influence on the performance of this method. The accuracy of the localization result only comes from a map of the monitoring space and the location information of some survey points.

An extension of the previous generic SVR method [56] was proposed by Wu et. al. [2], which contributed their efforts on solving the data missing problem in outdoor GSM networks. In this research, they analyzed the serious influence of the missing data problem to the accuracy of the SVR based mapping technique. During the field test, they observed that more than 80% of the record signal strength data is incomplete. This feature makes the utilization of SVR learning theory based approach in the outdoor GSM

38

environment a challenge problem. A missing value insensitive kernel, Sum of Exponential (SoE) kernel, was introduced in order to handle the missing value problem and to provide accurate regression results. Furthermore, a 1-norm loss function that is insensitive to small deviations is applied in the standard SVR method in order to build a linear regression model. The significant feature of the SVR approach is its ability to regress on the nonlinear expansion of the original input vector but not itself. This feature distinguished the SVR based method from other empirical methods. Unfortunately, this new method is still based on the assumption that the training data set is with time relevant feature. Thus, a novel training data set update method is necessary in the implementation of learning theory based methods in outdoor cellular networks.

Türkyılmaz [4] proposed a novel fingerprint technique, environment-aware RSS-based location estimation (EARBALE), which utilizes the artificial neural network (ANN) [64] based model to identify the radio propagation environment. This is achieved by using empirical data treated by preprocessing and dimensionality reduction via the decision tree (DT), and applying ANN based classification. After determining the environment feature, a matched deterministic Hata model [65] will be used, combined with the basis of the triangulation calculations. This is based on the observation that the radio propagation characteristic of different environments (urban area, suburban area, and open area) is significant and varied. The primary advantage of this method is its lightweight cost of integration onto any RSS parameter based location estimation scheme. The implementation of the EARBALE model only requires a small number of initiatory work and training processes.

Although EARBALE provides huge resource savings for location estimation in

cellular networks, the use of the propagation model after the ANN based classification is possible to introduce variances which will degrade the accuracy of the estimated location results.

2.3   Summary

This chapter presented some selected research work summaries that related to position related parameter extraction and position estimation methods used in cellular networks. It can be concluded from this chapter that although many machine learning theory based position estimation techniques have been presented, there is a scope of improvement in utilizing the hybrid TOA and SS parameters and BFA to further clean the EFs. Additionally, no technique exists in the literature that exploits the update of training data sets. A generic model can be designed to perform the functionality of position estimation and error reduction concurrently in outdoor cellular networks.

CHAPTER 3

CONGESTION REASON ANALYSIS BASED REAL-TIME ROUTE GUIDANCE SYSTEM FOR MULTI-LANE ROADS

3.1   Introduction

In this chapter, we developed a new navigation system for reducing travel time of an individual driver and distributing the flow of urban traffic efficiently in order to reduce the occurrence of congestion. Prediction based route guidance systems show great improvements in solving the inefficient diversion strategy problem by estimating future travel time when calculating accurate travel time is difficult. However, performances of prediction models that are based on the historical data set degrade sharply during a congestion situation.

In the metropolitan region, most congestion or traffic jams are caused by the uneven distribution of traffic flow that creates bottleneck points where the traffic volume exceeds the road capacity. Additionally, unexpected incidents are the next most probable cause of these bottleneck regions. Moreover, most drivers are driving based on their empirical experience without awareness of real-time traffic situations. This unintelligent traffic behavior can make the congestion problem worse.

Data fusion (DF) technique plays an inevitable role in intelligent transportation systems (ITS) with the objective of reducing transportation time and gasoline consumption, guaranteeing passenger safety, and avoiding congestion. This technique has contributed in specific tasks such as advanced traveler information systems (ATIS), advanced driver assistant (ADAS), traffic demand estimation, traffic forecasting and

monitoring, network control, automatic incident detection (AID), and crash analysis and prevention [66]. In this paper, we concentrate our efforts on the first scenario since there is an increasing requirement for congestion avoidance in urban areas.

Every year, the US wastes more than 10 billion gallons of fuel and 63 billion dollars due to traffic congestion [67]. Additionally, the annual travel delay that US citizens experience has increased fivefold within 20 years. In order to reduce the waste of time, fuel and money, the most important issue is to reduce the time spent in congestion during the trip. The shorter the time that people spend on the road, the less chance that they will be involved in congestion. The most straightforward method to achieve this is to build new roads or widen existing thoroughfares. However, construction of new roadways will cost millions of dollars and take long period of time, not to mention the possibility of causing more serious congestions during construction period. Thus, an intelligent route guidance system with accurate and reliable real-time traffic data would be a good option.

This section focuses on developing a navigation system which uses the AMPRFP algorithm to estimate accurate travel time of each arc based on both historical traffic behaviors and real-time non-linear affections caused by neighboring arcs' traffic patterns. The performance of this system was evaluated in a multi-lane urban area. The earlier work of authors [68] showed that the prediction based real-time fastest path (PRFP) algorithm can help drivers save travel time in single lane urban traffic networks, especially when they experienced frequent congestion in their trip. It also can improve the efficiency of urban traffic management by reducing the number of congestion events.

However, in the real world, most routes are with multiple lanes. And in this scenario, the cause of congestion cannot be identified simply by incidents due to the existence of

extra lanes that can distribute the traffic flow to bypass them. In [69], authors summarized that in the multi-lane urban traffic network, the congestion is mainly caused by the traffic volume exceeding the capacity of the arc. This is called "recurring congestion" which accounts for roughly half of the total congestion. The other type of congestion which is named "nonrecurring congestion" is caused by temporarily taking away part of the roadway. These new features were well considered in the building of the AMPRFP algorithm.

3.2   Route Guidance Techniques

Route guidance is the real-time process of guiding the driver along the route generated by a route planner. In general, vehicle navigation algorithms can be classified into two classes, the shortest path algorithm and the fastest path algorithm.

3.2.1   Shortest Path Algorithm

The first algorithm designed to provide optimal route guidance services is the shortest path algorithm [70]. It concerns route length as the impedance and tries to find the arc list with the shortest distance between the start vertex and the destination vertex. It was the best option in the last 20 years when the estimation of average speed and traffic flow of each arc was difficult. However, since the length of the arc is fixed, this system cannot respond to real-time changes in the route condition.

3.2.2   Fastest Path Algorithm

To overcome this problem, other form of users' information, the travel time, is used as a measure of cost for route selection strategies.

A. ADVANCE Program

One such kind of system is named ADVANCE program [71], which was an in-vehicle system providing route guidance in real-time. It operates in the northwestern portion and northwest suburbs of Chicago. In ADVANCE program, dynamic travel time information about expressways and local streets is used to determine the optimal routes for drivers. However, the estimation of travel time with high accuracy is a particularly difficult task especially when congestions or incidents happened. Thus, in ADVANCE program, arcs involved with congestions and incidents were treated as broken and were excluded from the arc data set. Obviously, this would introduce inefficient diversion strategy problem [72] into the selected optimal path. For instance, if a route will be blocked by an incident or congestion for 10 minutes, and the vehicle needs to take 15 minutes to reach it, diverting this vehicle to the nearby arc which is 20 minutes away is not efficient. Not to mention the increase of the possibility that it may be jammed by incidents in other paths.

B. Traffic Flow Forecasting

To overcome limitations of travel time estimation, the traffic flow forecasting approach has received incremental attention in the past years. Messer et. al [73] announced the first work on travel time forecasting. Later, a dynamic route guidance system (DRGS) [74] was proposed for estimating and predicting travel time and then weighted to get optimal routes. In order to further improve the performance of the prediction model, researchers have applied various techniques to this problem such as Bayesian framework [75], Linear regression [76], Kalman filtering [77], neural networks [78], and kernel estimator [79].

Currently, people predict the travel time based on the time-varying features related to traffic behaviors. It means the future travel time has some level of relationship with historical (daily, weekly, and seasonal) traffic patterns. Unfortunately, the fluctuation of traffic flow caused by congestions and incidents that happened in related arcs which are quite different from historical records is not well considered in these prediction methods.

3.3　System Architecture

The overall system architecture for the wireless real-time route guidance system is shown in Fig. 3.1. The system is designed to satisfy requirements of real-time traffic navigation at the fastest path level. There are two primary components in the system: 1) the mobile unit, and 2) the route guidance system server.
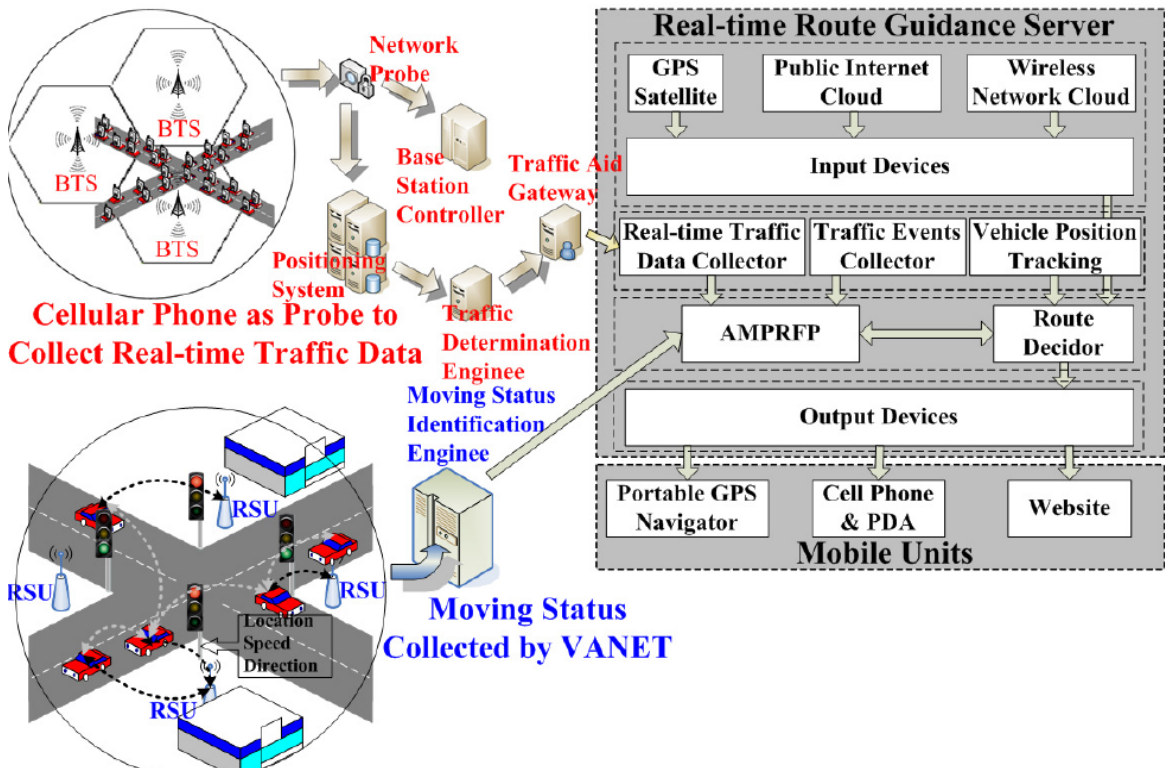

Figure 3.1: System Architecture

3.3.1   Mobile Unit

The mobile unit is a hand-held equipment which is integrated with a GPS signal receiver. Its main function is to decode the real-time data sent from the server. The unit can be a cell phone, a PDA, or a portable GPS navigator. The mobile device is designed to operate simple functions. Typically, mobile units cannot support high speed and complex computation. Hence, in the design of our system, using the browser/server model (Calculation is totally done at the server side; the presentation task is implemented at the mobile unit side.) becomes a good option to enable the mobile unit with the support of the real-time route guidance service. Additionally, in order to support multi-lane navigation service, the localization method applied by the mobile unit is the differential GPS (DGPS) [80] technique which can provide lane level position estimation service.

3.3.2   Route Guidance System Server

The guidance server can be a powerful server machine which is the most important part of this system. From Fig. 3.1, we can see that the server consists of four components: 1) the input module; 2) the data collection module; 3) the path calculation module; and 4) the output distribution module.

The input module is responsible for converting different types of input data to a standard and recognizable format for the system to process. The data collection module focuses on requesting real-time traffic data from the third party's database. The real-time data includes: 1) average velocity of monitored arcs; 2) accident reports of the urban networks; and 3) current location of customers. Based on these real-time traffic data, the path calculation module executing the AMPRFP algorithm first estimates the travel time

46

of each arc in the traffic network. After that, it calculates the near optimal fastest path for the current customer. Finally, the path information is converted into suitable data types for different mobile units. The data is sent through the output distribution module.

3.3.3   Real-Time Traffic Data Collection

Current techniques for real-time traffic information collection can be mainly divided into two categories: inductive loop sensors and probe based methods. Using inductive loop sensors, the speed can be estimated by comparing the distance between vehicles crossing them within a unit of time (60 second is the standard value used in the United States). This method can provide reliable real-time traffic data. However, the cost of building this network so expensive that the government has to sacrifice the secondary and tertiary roads, and focus on only some of the main highways and freeways. For instance, in Kansas City, the installation of a sensor system on a 70mile highway costs about $15 million. Thus, one possible way to solve this problem is to use cellular phones as probes to provide traffic information, since more than 80% of drivers will keep their cell phones open during the whole trip. This is achieved by routinely transmitting mobile phones' position information, in an anonymous format, back to the network to calculate traffic flow information using triangulation, cell sector statistics, or pattern matching algorithms [81]. Even when no voice connection is established, this process can still operate based on the characteristic of the cellular network. This is a good way to provide full coverage real-time traffic monitoring services in the urban city.

3.3.4   Moving Status Identification

In our AMPRFP algorithm, the detailed moving status of an individual vehicle, especially the leading vehicle, is critical to identifying the cause of congestion. For instance, without the knowledge of the moving direction of a leading vehicle which caused the congestion in current arc, the possible travel time of this arc cannot be calculated. However, the collection of this kind of information is difficult until the development of the Vehicular Ad-Hoc Network (VANET). The main goal of VANET is providing safety and entertainment information for passengers [82]. This is achieved by two components: the in-vehicle device and the roadside unit (RSU). The in-vehicle devices are some special electronic devices installed inside vehicles to provide Ad-Hoc Network connectivity. RSU acts as the wireless LAN access point to provide communications with other infrastructures. Vehicles involved in VANET will periodically broadcast their driving status (like driving direction, driving speed, current location, etc) through dedicated short-range communications (DSRC) channel. This one-way or two-way wireless channel is designed specifically for automotive use with 75MHz of spectrum in the 5.9GHz band in the USA [82]. This information will be received by associated vehicles and RSUs to support real-time services, such as intersection collision avoidance, road sign alarms, and in-place traffic view. Thus, as shown in Fig. 3.1, when the leading vehicle is close to the end of the current arc, its driving information can be collected by the RSU installed along the roadside and then sent to the route guidance server to construct traffic views of the whole system.

3.3.5   Congestion Detection

In three phase traffic theory [83], the congested traffic is defined as the situation when the speed of a vehicle is lower than the minimal speed in free flow. The threshold value of the minimal speed can be calculated using equation (3-1).

$$v_{min} = \frac{q_{max}}{k_{crit}}$$
(3-1)

In (3-1), q_{max} is the maximum flow rate (in vehicles per hour) of an arc that can be achieved in the free flow. Parameter k_{crit} is the corresponding critical density (in vehicles per mile). If the average velocity of an arc is higher than v_{min}, the traffic condition of this arc can be identified as free flow while in the reverse situation; this arc is detected as been congested.

In real urban networks, the v_{min} of each arc is varied. In order to standardize the definition of congestion and to reduce the complexity in the design of the AMPRFP algorithm, based on empirical research results, v_{min} in freeway is defined as 45mph [69]. In local arterial and city roads, the definition of v_{min} is

$$v_{min}^i = (0.8 \times 2 - (1 - \frac{k^i}{c^i})) \times v_{limit}^i$$
(3-2)

where k is the current density (in vehicles per mile) of arc i, c is the total capacity (in vehicles per mile) of i, and v^i_{limit} is the speed limit of i [84].

3.4   AMPRFP Algorithm

With the fast development of real-time traffic data monitoring technology introduced before (section 3.3.3), it is possible to receive accurate and on-time traffic information and maintain the overview of the whole urban traffic network. For instance, in traditional

time estimation techniques, the jammed or incident involved arcs were treated as broken and excluded from the route list temporarily. However, in the AMPRFP algorithm, those jammed arcs will not be simply excluded, but they will be assigned predicted travel time based on the analysis of the reason of congestion. This is the key idea of the design of this algorithm.

3.4.1   Related Research

In order to provide optimal routes for drivers to reduce their travel period, the chief concern is to identify a specific weight parameter that can be used for evaluation. In this research, we selected the estimated travel time of each arc in each time point (defined as every second) as this parameter. An efficient shortest cost algorithm is also important to the quality of the evaluation job. Additionally, in order to generate high quality travel time data, the careful selection of a prediction method is necessary and will be discussed later.

A. Shortest Cost Algorithm

[70] tests the performance of the most famous shortest path algorithms using real road networks data. From their conclusion, we learned that the Dijkstra's Approximate Buckets Algorithm (DIKBA) shows superior performance in solving one-to-one or one-to-some shortest cost problem. Dijkstra's algorithm was conceived and designed by Dutch scientist Edsger Dijkstra in 1959 [85]. It performs well when people try to solve single-source shortest path problems in a network without negative edges. The operation of DIKBA is shown below in pseudocode. In Algorithm 1, **V** and **E** represent the set of vertices and edges in the graph respectively. s is defined as the source node while $d[v]$ is

the tentative distance of *v*. (*v, u*) is the edge connecting vertex *v* and *u* and *d_between(v, u)*

shows the distance between them. Bucket *B* is an array that stores a group of vertices in

each element *B[i]*. Vertices in *B[i]* satisfy that $d[v] \in [i\Delta, (i + 1)\Delta]$ where Δ is a positive

real number that denotes the bucket width.

Algorithm 1 : DIKBA($\mathbf{V}, \mathbf{E}, s, \triangle$)

1: **for** each vertex v in **V** do
2:      heavy(v) := $\{(v, u) \in \mathbf{E} : d\_between(v, u) > \triangle\}$
3:      light(v) := $\{(v, u) \in \mathbf{E} : d\_between(v, u) \le \triangle\}$
4:      d(v) := ∞
5: **end for**
6: **if** $d[s] > 0$: **then**
7:      remove s from $B[\lfloor d[s]/\triangle \rfloor]$
8:      insert s into $B[0]$
9:      d[s] := 0
10: **end if**
11: i := 0
12: **while B** is not empty: **do**
13:      **S** := null
14:      **while** $B[i] \neq null$ **do**
15:          **REQ** := $\{(u, d[v] + d\_between(v, u)) : v \in$
              $B[i] \wedge (v, u) \in light(v)\}$
16:          **S** := **S** $\cup$ B[i]
17:          B[i] := null
18:          **for** each (v, x) $\in$ **REQ do**
19:              **if** $d[v] > x$: **then**
20:                  remove s from $B[\lfloor d[v]/\triangle \rfloor]$
21:                  insert s into $B[\lfloor x/\triangle \rfloor]$
22:                  d[v] := x
23:              **end if**
24:          **end for**
25:      **end while**
26:      **R** := $\{(u, d[v] + d\_between(v, u)) : v \in \mathbf{S} \wedge (v, u) \in$
          $heavy[v]\}$
27:      **for** each (v, x) $\in$ **REQ do**
28:          **if** $d[v] > x$: **then**
29:              remove s from $B[\lfloor d[v]/\triangle \rfloor]$
30:              insert s into $B[\lfloor x/\triangle \rfloor]$
31:              d[v] := x
32:          **end if**
33:      **end for**
34:      i := i + 1
35: **end while**
36: **for** each vertex v in **V** do
37:      return d[v]
38: **end for**

In our research, most drivers in the urban traffic network need to reach their destinations in the shortest possible time, and most people demand a single destination at each time. Thus, it is a one-to-one shortest cost problem and so DIKBA is an appropriate choice.

B. Prediction Model

In order to avoid traffic congestion with real-time and reliable traffic information, the key point is to predict and estimate the travel time of each arc. In 1973, Messer, Dudek and Friebele [73] announced the possibility of time prediction in their work. After that, many researchers [75-79] have concentrated their efforts on improving the accuracy of the predicted time.

SVR

Support vector regression (SVR) is one of the most important applications of the support vector machine (SVM) theory and was developed by Vapnik [61, 62]. The core concept of SVR is its general belief that the divisibility of data can be more distinct after transferring data to a high-dimensional feature space [62]. The advantage of SVR, that makes it superior than conventional empirical methods, is its ability to minimize the upper bound of the expected risk to improve significantly the generalized performance of regression.

Pioneering work by Wu et al. [79] found out that the computation of travel time has a tight relationship with vehicle speed, traffic flow, occupancy, weather condition, and traffic incidents. All of these increase the complexity and difficulty of accurate travel time prediction. The key idea behind the SVR algorithm based travel time prediction method is that it considers the future value of travel time correlating with the time-variant

historical data set. In [86], travel time is split into three categories: historical, current, and predictive. There are linear relationships existing between them, which are shown in Fig. 3.2.
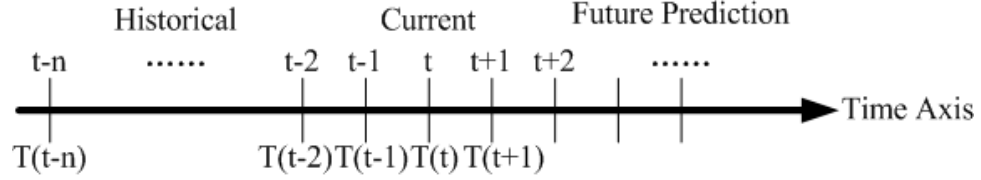

Figure 3.2: Travel Time Relationship

The relationship between the predicted travel time and the training data can be demonstrated using equation (3-3). Suppose there is a training set of travel time $\mathbf{T}_i \subset \mathbb{R}^n, i = 1, ...n$ where $\mathbf{T}_i = \{t_i\}_{i=1}^n$ and n is the number of training samples. As defined in SVR, there is a mapping function ϕ that can transfer the input space $\mathbb{R}^n$ into a high-dimensional feature space **F**. We need to build a linear function $f()$, which can model each transferred data $\phi(t_i)$ to its associated travel time period p_i. The predicted travel time period p^* at the future time point t^* is:

$$p^* = f(\phi(t^*)) = \mathbf{w}\phi(t^*) + b\mathbb{R}_n \tag{3-3}$$

In order to get the result from (3-3), we need to decide the vector w and the offset scalar $b \in \mathbb{R}$. This is achieved during the learning process. First we define the ϵ-intensive loss function l as:

$$l = \begin{cases} 0, & if\ |p_i - f(\phi(t_i))| < \epsilon \\ |p_i - f(\phi(t_i))| - \epsilon, & otherwise \end{cases} \tag{3-4}$$

Since ϵ cannot always cover otherwise infeasible limitations of the convex optimization problem, slack variables ξ_i and $\tilde{\xi}_i$ are applied to simplify l.

$$l = \xi_i + \tilde{\xi}_i \tag{3-5}$$

Additionally, the linear function $f()$ needs to satisfy Mercer's condition [62], which means the training process needs to solve following optimization problems:

$$\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n}(\xi_i + \hat{\xi}_i) \tag{3-6}$$

$$\begin{cases} p_i - \mathbf{w}\phi(t_i) - b \le \epsilon + \xi_i, \\ \mathbf{w}\phi(t_i) + b - p_i \le \epsilon + \hat{\xi}_i. \end{cases}$$

$$(\xi_i, \hat{\xi}_i, \epsilon, C \ge 0, i = 1,2,\dots,n). \tag{3-7}$$

The new introduced parameter C is used to balance the flatness of the linear function and its ability to tolerate deviations. The value of C is assigned before the training phase. In order to achieve a better balance, we set C as 1000. Next, after maximizing the corresponding dual problem [62], $\mathbf{w}$ can be solved in equation (3-8).

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\,\phi(t_i), \alpha_i, \alpha_i^* \in [0,C] \tag{3-8}$$

In (3-8), two new parameters are introduced: α_i is the lagrange multiplier and α_i^* is its estimation. Bring (3-8) to (3-1), we have:

$$\begin{cases} p_i = f(\phi(t_i)) = \sum_{j=1}^{n}(\alpha_j - \alpha_j^*)k(t_i, t_j) + b \\ \sum_{j=1}^{n}(\alpha_j - \alpha_j^*) = 0, \alpha_j, \alpha_j^* \in (0,C) \end{cases} \tag{3-9}$$

where $k()$ is the kernel function used to determine the quality of the space transfer and to simplify the calculation. The variable b can be calculated based on the Karush-Kuhn-Tucker (KKT) conditions [62] shown in (3-10).

$$b = \begin{cases} p_i - (\mathbf{w}, t_i) - \epsilon & \alpha_i \in (0,C), \alpha_i^* = 0, \\ p_i - (\mathbf{w}, t_i) + \epsilon & \alpha_i^* \in (0,C), \alpha_i = 0. \end{cases} \tag{3-10}$$

Based on the result of [77, 79], it can be concluded that SVR is most suitable for this system. And in this research, the mySVM software kit [87] was chosen to operate all the SVR experiments.
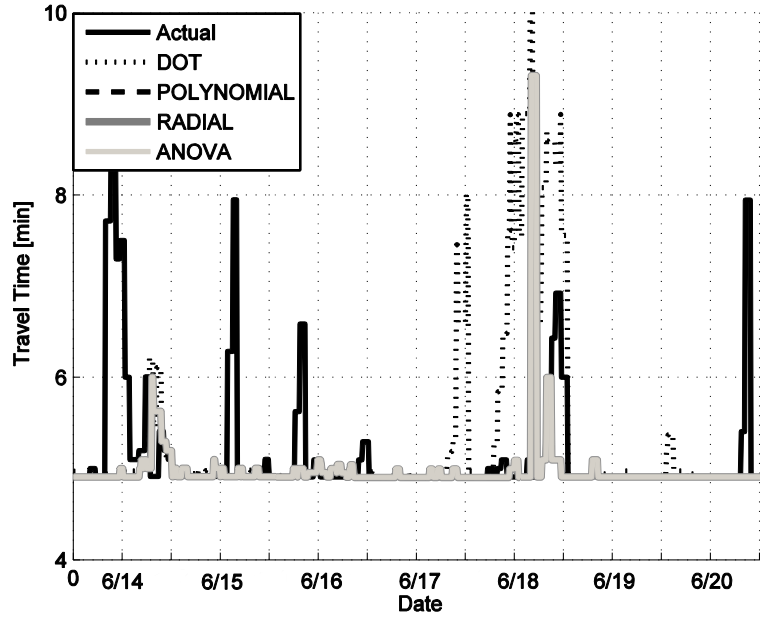
Kernel Function

As a kernel based mapping algorithm, SVR can augment the training data to a high-dimensional feature space using the kernel function. During the development of SVR theory, different kernels were designed to be applied in different applications. In previous research [68], we only considered the performance of SVR with the dot kernel, which is the simplest one to be integrated, but has a lower accuracy level and less ability to deal with missing values. Thus, in the new experiment, we compared the performance of other kernels shown in Table 3.1, such as the polynomial kernel, the Radial Basis Function (RBF) kernel, and the Analysis of Variance (ANOVA) kernel, using the field test data without congestion. The detail of field tests can be seen in following section.

Table 3.1: Kernel Functions

Kernel	Function	Parameter
Dot	$x * y$	
Polynomial	$(x * y + 1)^D$	D
Radial	$exp\,(-\gamma\|x - y\|^2)$	γ
ANOVA	$(\sum_{i=1}^{l} exp\,(-\gamma(x_i - y_i)))^D$	D, γ

[After trying several combinations of and D, we finally concluded that set D=3 and γ=0.3 can achieve the best prediction result.]

In Fig. 3.3(a), the ability to provide accurate prediction results of these four kernel functions is evaluated by comparing the predicted travel time with the measured travel time at the same time point for the same arc and same direction. From this figure, we can learn that the polynomial kernel provides the most dynamic prediction results. However, it is difficult to identify which kernel is the best due to performance variance. Therefore, two statistical methods, relative mean errors (RME) and root-mean-squared errors (RMSE) are introduced to further compare their performances.

(a) original value comparison
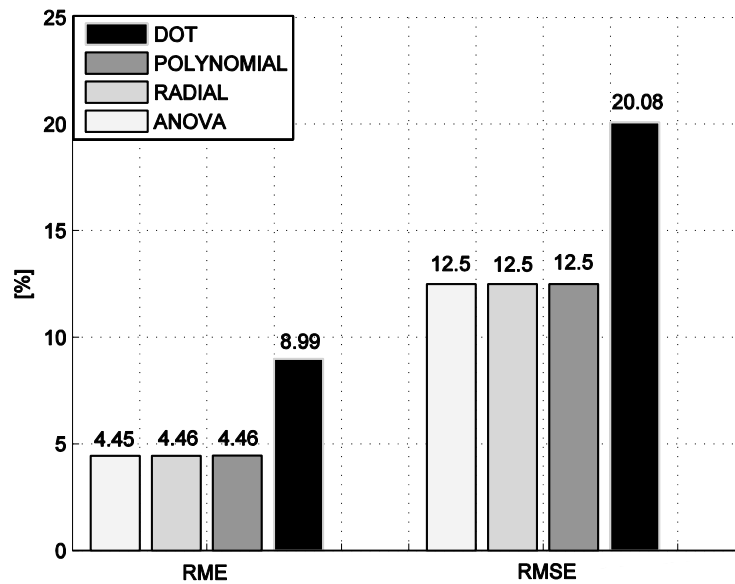
(b) RME & RMSE for different kernel

Figure 3.3: Prediction performance of different kernel functions. (a), (b).

In (3-11) and (3-12), n is the size of time points sampled in the experiment. $T_{i,j}$ is the

actual travel time of arc j at time point i and $T_{i,j}^{*}$ is the predicted travel time at the same arc and at the time point i.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_{i,j}-T_{i,j}^{*}}{T_{i,j}}\right|,$$ (3-11)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_{i,j}-T_{i,j}^{*}}{T_{i,j}}\right|^{2}}.$$ (3-12)

In (3-11) and (3-12), n is the size of time points sampled in the experiment. $T_{i,j}$ is the actual travel time of arc j at time point i and $T_{i,j}^{*}$ is the predicted travel time at the same arc and at the time point i.

After evaluating the result in Fig. 3.3(b), we conclude that the ANOVA kernel is able to provide the most accurate prediction results. The RME and RMSE values of using the ANOVA kernel are about 4.45% and 12.5% respectively. These results are close to the simulation result of [79], which is also based on real highway traffic data.

3.4.2   AMPRFP Algorithm

With the fast development of real-time traffic data monitoring technology introduced before (section 3.3.3), it is possible to receive accurate and on-time traffic information and maintain the overview of the whole urban traffic network. For instance, in traditional time estimation techniques, the jammed or incident involved arcs were treated as broken and excluded from the route list temporarily. However, in the AMPRFP algorithm, those jammed arcs will not be simply excluded, but they will be assigned predicted travel time based on the analysis of the reason of congestion. This is the key idea of the design of this algorithm.

A. Traffic Pattern Classification

Based on three phase traffic theory [83], we divided the route condition into four categories.

Free Flow Phase

The first category is the free flow phase. This is the situation when there are vehicles (the traffic density $\rho > 0$) moving normally with the average speed (v_{avg}) in an arc, and this average speed is larger than the minimum threshold of the traffic congestion level (v_{min}). The travel time period (p) can be estimated using:

$$p = \frac{l}{v_{avg}} * 3600, \tag{3-13}$$

$$v_{avg} = \frac{\sum_{i=1}^{n}(\prod_{j=1, j \neq i}^{n} v_i v_j)}{n \prod_{k=1}^{n} v_k}. \tag{3-14}$$

where l is the length of the selected arc.

Empty Flow Phase

Different from three phase traffic theory, we add an additional category, the empty flow phase, which is different from the free flow phase because there is no vehicle in the arc ($\rho = 0$) and so no average speed can be detected. Thus, the same equation (3-13) is used by replacing only v_{avg} with the maximum speed (or speed limit: v_{limit}) of this arc.

Synchronized Flow Phase

Third, the synchronized flow phase is defined as the situation where congestions happen in the arc. In this situation, as analyzed in section 3.3.5, the traffic speeds in both lanes are very slow ($\rho > 0$ and $0 < v_{avg} < v_{limit}$), and the travel time is difficult to estimate. Additionally, the velocities in neighboring lanes have the synchronization feature due to the lane changing behavior of drivers.

Traffic Jam

Fourth case is that of a traffic jam happening in the arc. All vehicles in the arc cannot move ($v_{avg} \approx 0$). The road density is reaching the maximum value ($\rho = \rho_{max}$).

Considering the above categories, it is obvious that the accuracy of travel time estimation drops significantly during the synchronized flow phase and the traffic jam phase. Thus, in the next part, we introduce a congestion reason analysis method to approximately calculate the travel time during these two phases.

B. Multi-Lane Congestion Analysis

In the congestion situation, factors that need to be considered include the quality of time-variant historical data, moving pattern of current arc, and the real-time traffic condition of neighboring arcs. These factors will influence the accuracy of the estimated travel time. After analyzing the characteristics of traffic patterns during congestion in the multi-lane urban traffic network, we classify reasons for congestion into following types.

Reason I: Incident

In multi-lane road systems, most incidents blocked only part of the arc but not all of it. Thus, the traffic volume rose because of the temporary disruptions in the arc and caused the sharp reduction of the travel speed. Such kinds of incident include: a broken vehicle, construction, weather, a crash, etc. The traffic pattern is shown in Fig. 3.4. In this situation, in order to calculate the possible travel time at time instance t, we need to consider the releasing time of the incident. This is due to the assumption that the congestion can be released after the incident is removed. After the incident is cleaned, the capacity of the current arc can be recovered and the traffic pattern will changed to the

free flow phase. However, the average speed during that time cannot be estimated. Therefore, the formula for travel time estimation can be defined as (3-15):
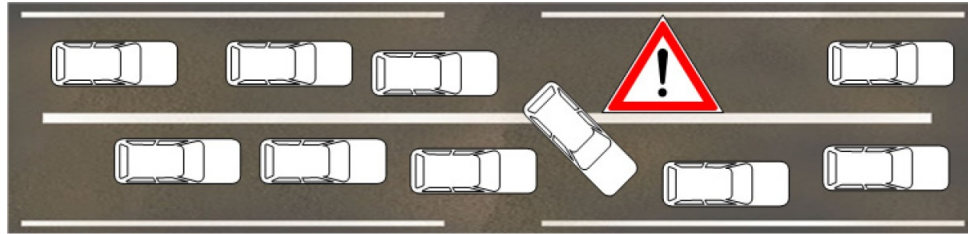
Figure 3.4: Incident caused road congestion.

$$p = \begin{cases} \dfrac{l}{v_{avg}} * 3600, & \dfrac{l}{v_{avg}} \leq p_{incident} \\ p_{incident} + SVR(t + p_{incident}), & \dfrac{l}{v_{avg}} \geq p_{incident} \end{cases} \qquad (3\text{-}15)$$

In the other situation, if there is just a single lane or the incident blocked the whole arc, shown in Fig. 3.5, the travel time of this arc is relied on when the incident can be released. Thus the travel time is measured approximately using the maximum time period required to solve this event. Using approximate value is based on the assumption that the time need to solve the incidents is much longer than the time for a vehicle to drive through this route. Thus it is acceptable to ignore the travel time but consider the time period used to solve the incident only.

Multi-lane Incident

Single lane Incident
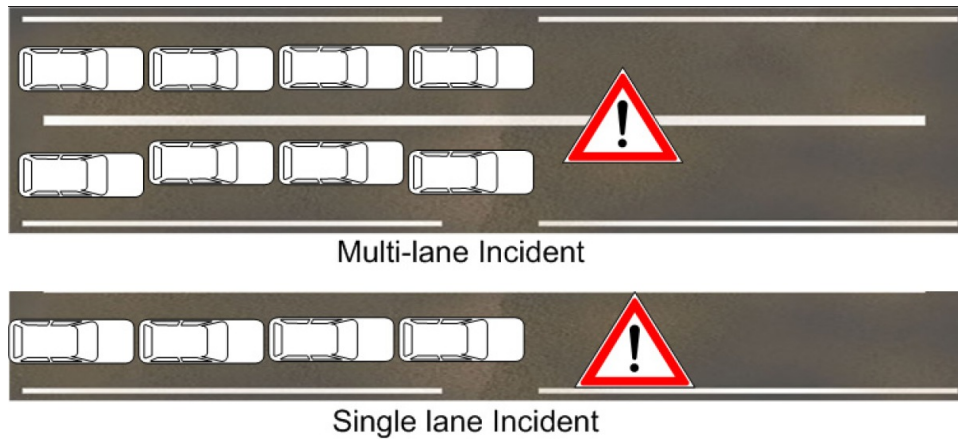
Figure 3.5: Incident caused road jam.

Reason II: Change of road architecture
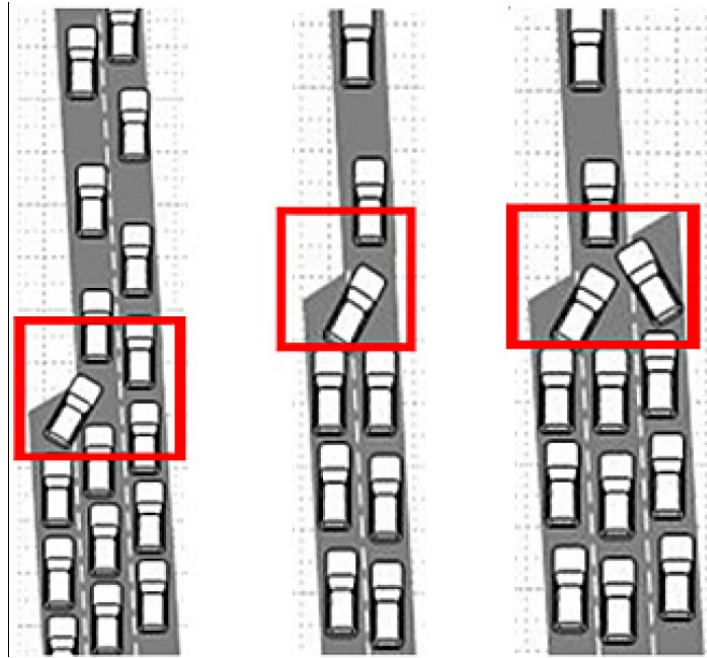
Figure 3.6: Merge lane caused road congestion.

The changing of route architecture, which reduces the road's capacity, may also be an important reason for congestion. This happens when a lane merge exists in the connection of two arcs or during part of an arc. The situation can be identified in Fig. 3.6. Unfortunately, in this situation, the congestion will remain until the number of vehicles in this arc is reduced and reach the minimum density of the two architectures. To calculate the travel time in this situation, the following formula can be used:

$$p = \frac{l}{v_{avg}} * 3600. \tag{3-16}$$

Reason III: Recurring

Sometimes there is no significant cause of congestion. In high density traffic, even a tiny disturbance such as getting too close to the other vehicle or braking too fast, can be quickly amplified into self-sustaining traffic congestion. For instance, sometimes when

you are driving on a highway and the traffic suddenly slows down, you inch along for a long period, thinking that there must be some accidents that caused this congestion. However, suddenly all the cars ahead of you take off at high speed and there is nothing in front. Meanwhile, the cars behind you are still congested.
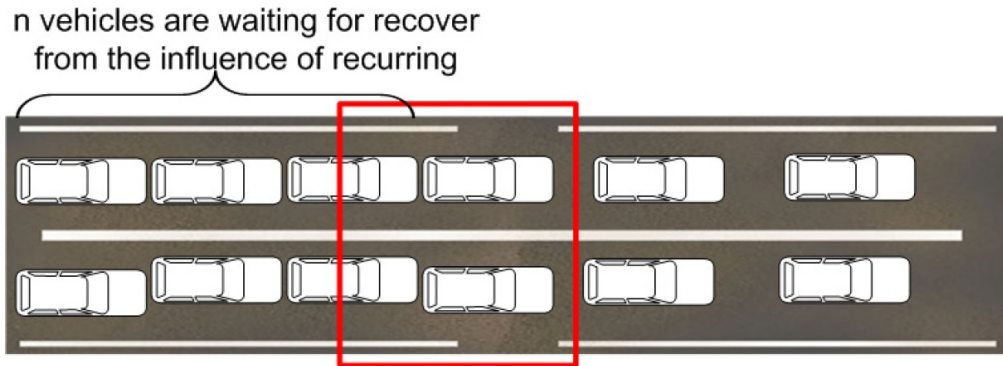
Figure 3.7: Recurring caused road congestion.

In this case, generally, a wave spread phenomenon can be observed. The congestion is spreading to the reverse direction of the current arc. A specific feature of recurring congestion is that at any time instance, we can identify the top vehicle of the congested queue. Ahead of this point, vehicles are accelerating to the normal speed, which is larger than v_{limit}. Based on this phenomenon, the travel time in this arc can be estimated based on the length of the congested queue. Suppose there are *n* vehicles blocked behind the congested point, shown in Fig. 3.7, the travel time of this arc at current time point can be calculated in (3-17):

$$p = n * \frac{l_{queue}}{v_{cong}} * 3600 + SVR\left(t + n * \frac{l_{queue}}{v_{cong}} * 3600\right). \tag{3-17}$$

C. Incident Detection & Clearance

Incidents, as one of the most important reasons for congestion [69], have great effect on the accuracy of travel time prediction and estimation. Generally, the source of

incidents that will cause travel time delay can be divided into two groups: planned events and unplanned events. Planned events are those actions that have a specific plan of execution time, position, and duration. For instance, the construction and police actions are two typical types of such events. Since the occurrence time and event duration are known in advance, we can use these data directly in the AMPRFP algorithm to decide $p_{incident}$. However, for unplanned events, such as a vehicle collision or an obstruction, it is difficult to decide their occurrence time and duration for clearance. In this system, we detect this kinds of incident rely on accident reports. From [88], we learned that the maximum period for clearing an incident is 90 minutes due to the implementation of the Rapid Incident Scene Clearance (RISC) program. This threshold value combined with reported occurrence time of each incident is used to calculate the approximate incident's duration $p_{incident}$.

D. Traffic Light Control

Traffic light control is a critical technique in maintaining the smooth movement of traffic flow and protecting the safety of pedestrians and drivers when they cross intersections. It also has an impact on travel time prediction and estimation. The traditional method to control the period of traffic light is called fixed time control. In this method, a timer is used to ensure each phase of the signal continues for a constant duration and repeats regardless of real time traffic situations. In order to improve the efficiency of the traffic light in controlling intersection traffic flow, the dynamic control method, with the support of computer and detector technologies, has become popular in current urban traffic networks. The dynamic control system can adjust timing and phasing of a signal to meet the change of traffic conditions in order to minimize the delay

of vehicles passing the intersection. In this research, we focus on forecasting the travel time of arcs that involved in congestion due to various reasons. Therefore, the dynamic control method is not appropriate due to the need of also predicting the future traffic condition of arcs before determining the signal's duration. It will introduce extra error because of the double prediction. Thus, in this research, we applied the fixed time control strategy in the design of the AMPRFP algorithm and the synthetic experiment. Based on [89], the average signal cycle length in Florida is 2 min. This value is incorporated with the iterative travel time estimation method to decide the travel time.

E. Iterative Travel Time Estimation
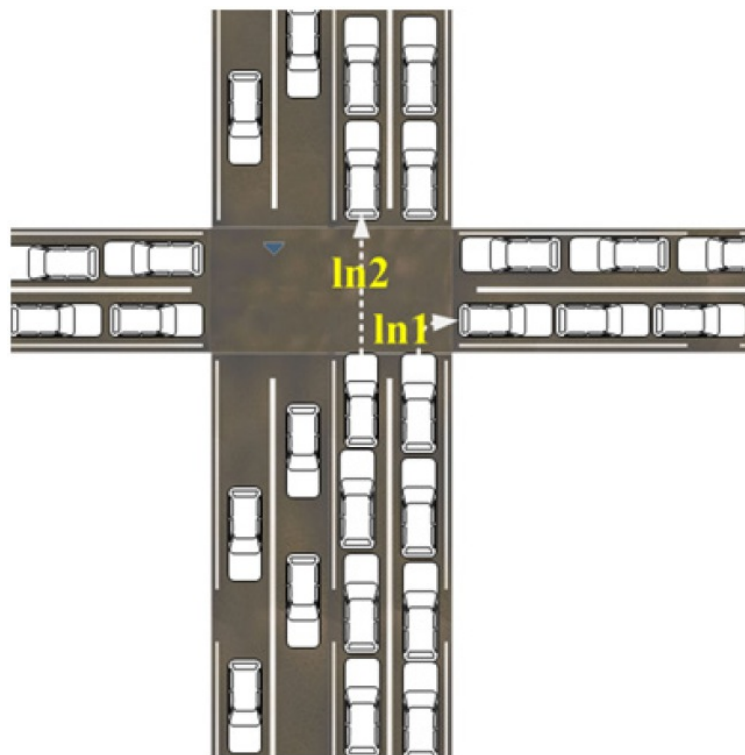

Figure 3.8: Real world example of a congested route.

When there is no incident or recurring happening in the arc, congestions and jams are generally caused by vehicles in the head of the arc which cannot move to their desired

arcs. One real world example that can illustrate this situation is shown in Fig. 3.8. Assume at time t, car2 in lane II ($ln2$) demands to go straight and enter the arc in front. While at the same time, car1 in lane I ($ln1$) needs to turn right to the right arc. Unfortunately, both the front arc and the right arc are experiencing congestion for some reasons (described in above part). Thus, the travel time of the arc where car1 and car2 are located is hard to estimate because it strongly relates to the time when car1 or car2 can recover from the jam. Either one of them can drive forward; vehicles in the lane can continue moving and even vehicles in the other lane can move by changing their lane. Suppose the travel time of two lanes where car1 and car2 are headed at time t are p_{ln1}^* and p_{ln2}^*. They can be estimated using equation (3-13, 3-15, 3-16, or 3-17). Then equation (3-18) and (3-19) can be used to calculate the travel time of p.

$$p_{lni} = p_{lni}^* + SVR(t + p_{lni}^*),$$ (3-18)

$$p = min_{1 \leq i \leq n}\{p_{lni}\}.$$ (3-19)

F. Fastest Path Selection

Finally, after updating the travel time for every arc, the fastest path can be calculated using the DIKBA. In this way, we not only reduce the travel time by guiding drivers to avoid most of the incidents and traffic congestions, but also by alleviating inefficient diversion strategy problems [72].

3.5 Experiment Setup

The simulation environment of the previous paper [68] was randomly chosen from the available data. It came from a small city, which was not convincing for representing the congestion problem in the big city. Thus, in the design of this experiment, we

carefully chose one of the busiest regions in Jacksonville to be the simulation environment. As the 10th biggest city in the United States, there are more than 1 million residents living and working in Jacksonville. Also, based on the data of [90], Jacksonville is the 49th most congested metro city all over the country. Thus, setting our test area in Jacksonville can make our conclusion adaptable to the big city scenario.
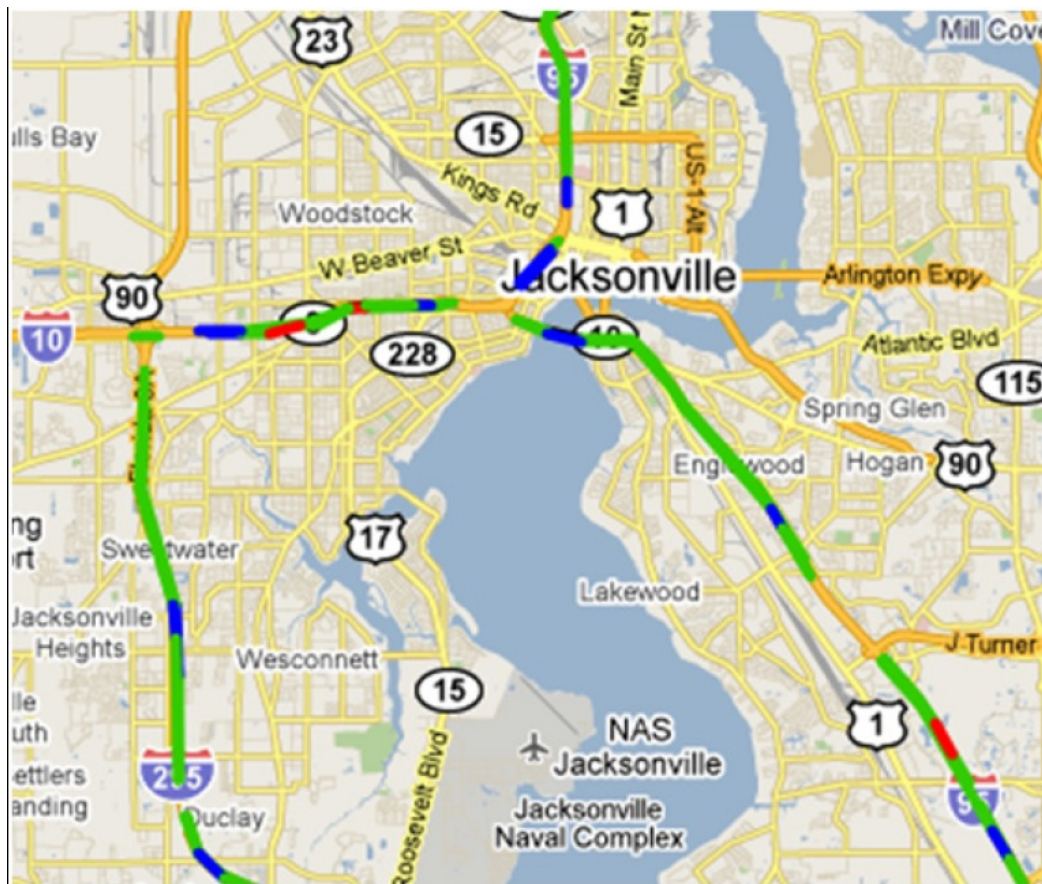
3.5.1   Data Preparation

Figure 3.9: Field Test Area from FL511.COM.

The experiment includes two sections: the field test and the synthetic experiment. In the field test, we focus on evaluating the ability of our AMPRFP algorithm in predicting accurate travel time during a congestion situation. The traffic data is provided by the

Florida Department of Transportation (FDOT). On their website, shown in Fig. 3.9, they constantly update the collected average speed from loop detectors that are deployed at I-95, I-10, and I-295 Highways. This data is updated once every 1 minute. Due to various reasons, the traffic data may be missed or corrupted. Thus, after analyzing the two month data collected from FDOT (from May 23rd, 2010 to June 29th, 2010), we choose a better portion of the data set between May 24th, 2010 and June 20th, 2010. During this period, there are no data missing. The data from the first three weeks are used as the training set, while the data of last 7 days are used as the testing set. By comparing the average speed data and the incident reports during the test period, we select three congestion events that the reason of congestion can be decided based on the available information shown in Table 3.2.

Table 3.2: Congestion Information

Event	Location	Reason
Congestion #1	I-95 SB from I-10 to Butler Blvd	Planned construction in Duval on I-95 south at Fuller Warren, 2 left lanes blocked.
Congestion #2	I-95 SB from I-10 to Butler Blvd	Crash in Duval on I-95 north beyond Butler.
Congestion #3	I-10 EB from I-295 to I-95	Crash in Duval on I-10 east at US-17, center lane blocked.

In order to test the performance of our system in local arterial, a synthetic experiment was designed. In the simulation, we selected one of the most congested areas in Jacksonville based on the National Traffic Scorecard 2008 Annual Report [91]. As shown in Fig. 3.10, the area is formed by 118 arcs and 86 vertices. The maximum length of these arcs is 0.19 mi and the shortest length is 0.002 mi. The maximum number of lanes for each direction is 3. In order to make the simulation result more distinctive, we created a

worse scenario where 60% of arcs in this area are with a traffic volume close to their maximum capacity. The direction, destination, location, and other related information of each vehicle and the incident's schedule for this area are generated randomly. Additionally, the total number of accidents, constructions, and so on is set to be a little higher than the historical maximum.
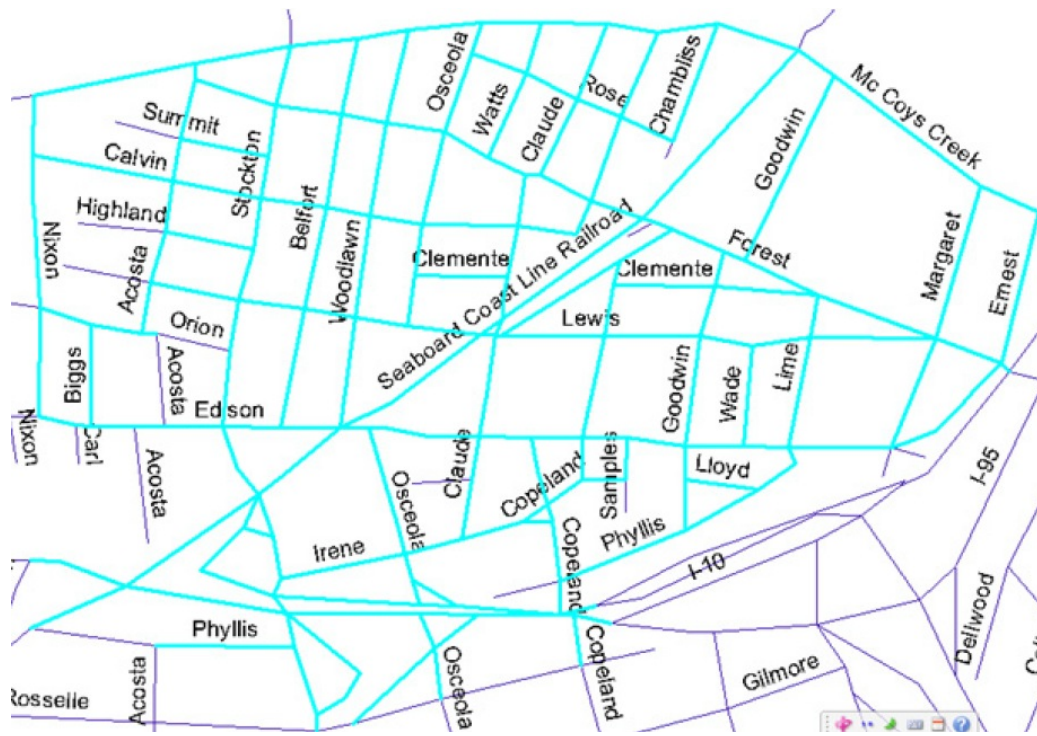
Figure 3.10: Simulation Area Map.

A famous microscopic traffic simulation software, namely TranSims, was used to operate the simulation. A C# program was developed to corporate with TranSims by calculating route lists for sampled vehicles using the AMPRFP algorithm, the shortest path algorithm, and the TFP algorithm. The default fixed-time control strategy of traffic lights in TranSims was selected to simplify the process of the simulation. The implementation of ANOVA based SVR was done using a mature software for SVM

research named mySVM [87]. Parameters were selected relying on empirically favorable values: $C = 1,000; \epsilon = 0.01; D = 3; \gamma = 0.3.$

3.5.2   Comparison of algorithms

In order to better present the performance of our system, some comparisons are necessary. In the simulation, the performance of the AMPRFP algorithm was compared with the traditional shortest path algorithm and the travel time based fastest path (TFP) algorithm.

A. Traditional Shortest Path Algorithm

As the classical route guidance algorithm, the shortest path algorithm plays a very important role in the navigation area [70]. It is very famous and has had large market share over a long period. The reason it is so popular is that it matches people's daily behavior. Most people always try to reach their destination by following the approximate shortest route learned from their experience. This is normal driving behavior. Also, in good traffic conditions, the shortest path can help them save money and time.

B. TFP Algorithm

The TFP algorithm, which is originated from the ADVANCE program [71], tries to search the entire road network and look for a path which has the shortest travel time to finish. It is helpful when used to decide the highest quality route. Hence, it can explore alternative paths efficiently to provide the optimal result for complex routing challenges. In the simulation, the TFP algorithm was abstracted to one suitable for the simulation while keeping most of its navigation features.

3.6   System Evaluation Results

In this section, we analyze the results of the field test and the synthetic experiment in order to demonstrate the effectiveness and superiority of our proposed AMPRFP algorithm.

3.6.1   Simulation Experiment I (Travel Time Estimation with Congestion)

As shown in section 3.4.1 (B), we found that the SVR prediction algorithm can achieve high accuracy when no congestion and no incident happened. However, the main goal of this system is to improve the ability of guiding drivers to save travel time without increasing the burden of the global traffic condition of the urban traffic network. This requires accurate travel time estimation when congestions and incidents happened.
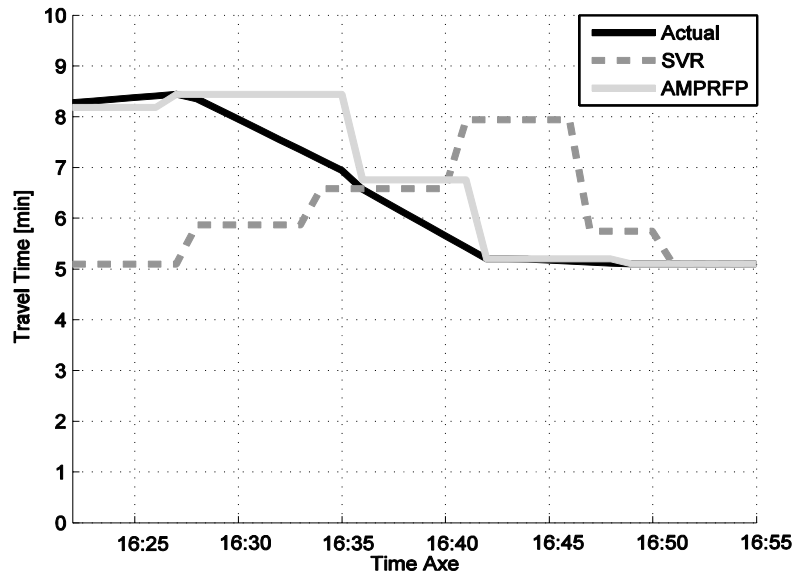
(a) Travel time estimation results for different algorithms for congestion #1
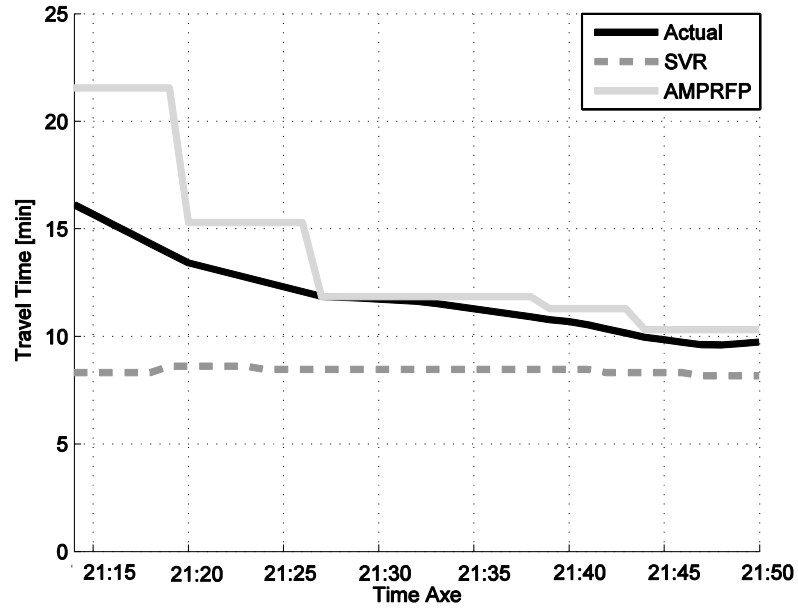
(b) Travel time estimation results for different algorithms for congestion #2
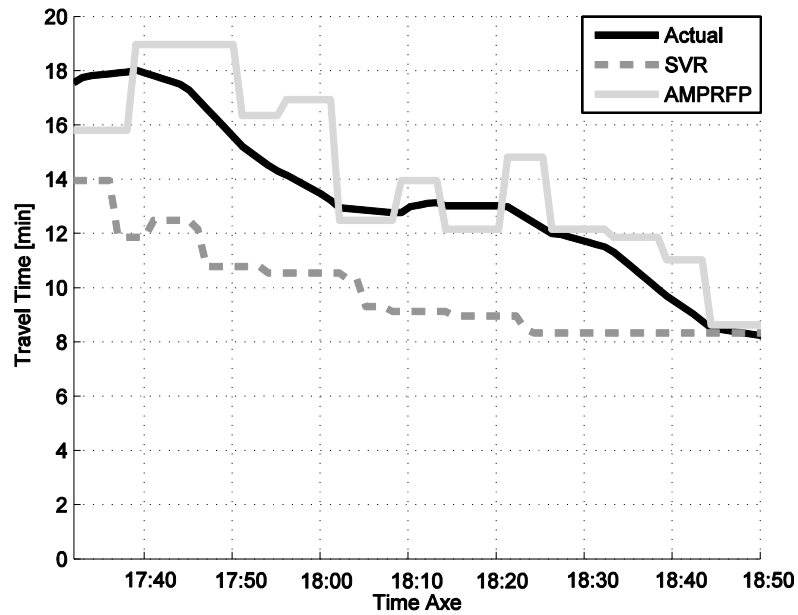
(c) Travel time estimation results for different algorithms for congestion #3

Figure 3.11: Comparison of travel time difference using different travel time

estimation algorithms (a), (b), (c).

In this experiment, we compare the performance of the SVR prediction algorithm and our AMPRFP algorithm in estimating travel time along congestion and incident situation. During the testing period, we selected three congestions, which happened on I-95 SB from I-10 to Butler Blvd and I-10 EB from I-295 to I-95 on different days of this week. The causes of these congestions were able to be decided by incidents that happened on the arc, or the adjacent arc, which was also monitored by the FDOT.

The experiment results, involve the predicted travel time over short distances (4.5 miles and 7.9 miles), are shown in Fig. 3.11. As expected, the SVR algorithm cannot reflect the sudden change of traffic patterns that are quite different from the past average. Admittedly, there is a time-lagged phenomenon for AMPRFP algorithm at the beginning of each change because of the impossibility of accurate prediction of the occurrence of incidents. Nevertheless, the curve of the AMPRFP algorithm follows the change of the actual travel time and has less predicted errors. The results in Table 3.3 show the RME and RMSE of these two algorithms. They demonstrate that the AMPRFP algorithm is able to reduce both RME and RMSE for at least 50% of those achieved by SVR algorithm for all three congested situations.

Table 3.3: Predicted Results in RME and RMSE of Different Algorithms

RME	SVR [%]	AMPRFP [%]
Congestion #1	27.34	14.3
Congestion #2	24.38	10.17
Congestion #3	23.56	5.21

RMSE	SVR [%]	AMPRFP [%]
Congestion #1	28.98	20.65
Congestion #2	26.13	12.24
Congestion #3	29.99	8.81

3.6.2 Simulation Experiment II (Ability for Reducing Travel Time)

In this experiment, we randomly take 10% of the sample vehicles from the whole car list. The path searching algorithm is set as the shortest path algorithm, the TFP algorithm, and the AMPRFP algorithm respectively. To make the result more rigid, all tests have the same initial settings.

Fig. 12 shows the comparison of the time difference between the shortest path algorithm and the TFP algorithm, the shortest path algorithm and the AMPRFP algorithm, and the TFP algorithm and the AMPRFP algorithm respectively. The time difference of car i following three algorithms is set as $t_{TFP}(i) - t_{shortestpath}(i)$, $t_{AMPRFP}(i) - t_{shortestpath}(i)$, and $t_{AMPRFP}(i) - t_{TFP}(i)$ respectively. Therefore, in Figs. 12(a),(b), vehicles guided by the TFP algorithm and the AMPRFP algorithm show more efficient performance than vehicles guided by shortest path algorithm due to their ability in avoiding congestion rely on real time traffic data.
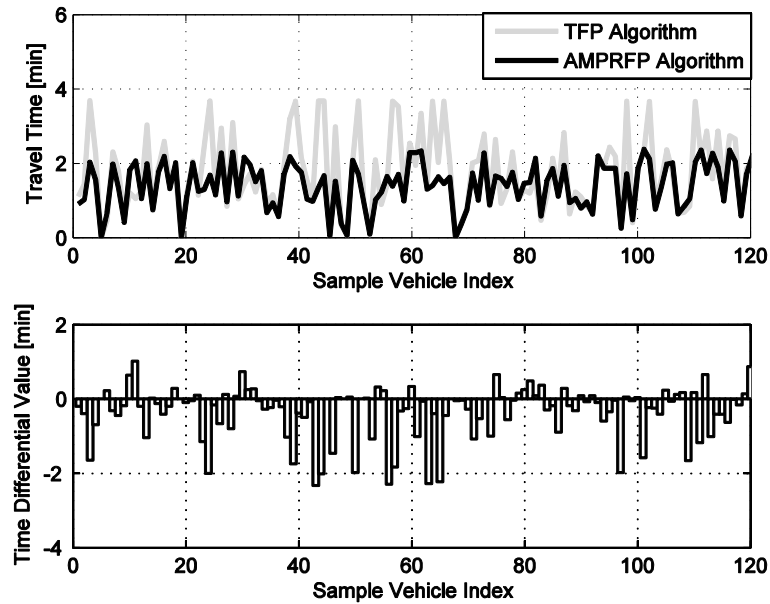
(a) Shortest Path Algorithm vs. TFP Algorithm
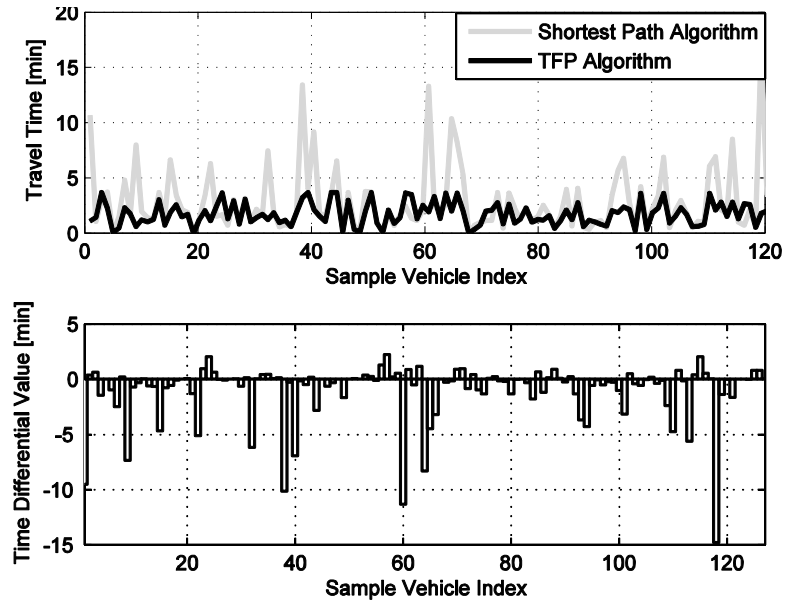
(b) Shortest Path Algorithm vs. AMPRFP Algorithm
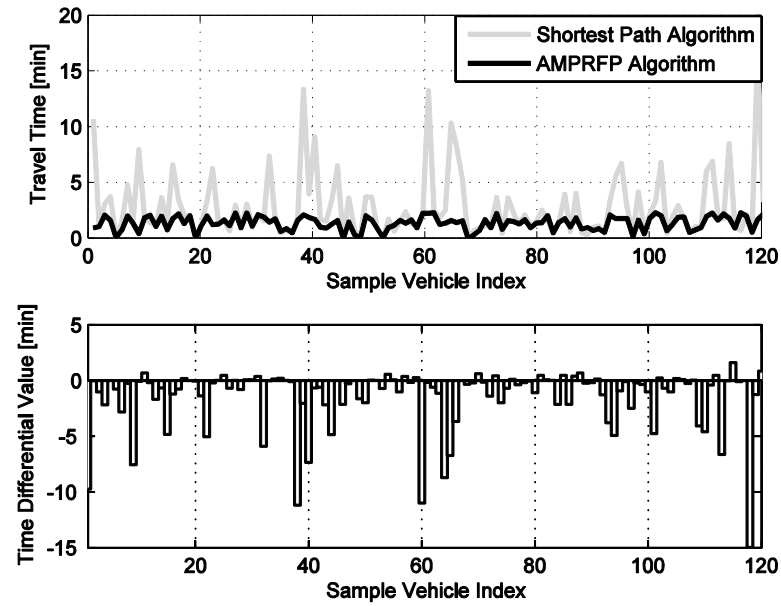
(c) TFP Algorithm vs. AMPRFP Algorithm

Figure 3.12: Comparison of travel time difference using different route guidance

algorithms (a), (b), (c).

In order to further analyze the performance of the TFP algorithm and the AMPRFP

algorithm since they both override the shortest path algorithm, in Fig. 12(c), the results of these two methods are compared. It is clear that the TFP algorithm has fewer opportunities to save travel time for drivers due to the inefficient diversion strategy problem.

Table 3.4 Comparison of the Ability to Reduce Travel Time Using Different Guidance Algorithm

Algorithm	Improvement Ratio [%]	Max Reduced Time [min]
Shortest Path Algorithm	40.94	2.20
TFP Algorithm	59.06	14.80

Algorithm	Improvement Ratio [%]	Max Reduced Time [min]
Shortest Path Algorithm	29.13	0.68
AMPRFP Algorithm	70.87	14.97

Algorithm	Improvement Ratio [%]	Max Reduced Time [min]
TFP Algorithm	30.71	1.02
AMPRFP Algorithm	69.29	2.33

Additionally, shown in Table 3.4, as expected, the sampled vehicles using the AMPRFP algorithm have a 70.87% and a 69.29% chances to reduce their travel time compared with other two algorithms. The data also demonstrates that compared to the shortest path algorithm, using the AMPRFP algorithm can improve the performance of the guidance system by about 2.3 times. By comparison, the TFP algorithm can only improve the performance of the guidance system by 1.5 times.

3.6.3　Simulation Experiment III (Urban Network Congestion Management)

The global optimization problem is also a very important parameter for estimating the quality of the proposed route guidance system. In the urban traffic network, most arcs

are connected with each other. Therefore, the change of traffic volume in one arc will also influence the traffic situation in others. A guidance system is not efficient if it can only solve the congestion problem of one driver but cause more congestions in other areas of the network. Therefore, in this case, we test the ability of congestion management for guidance systems that applying the TFP algorithm and the AMPRFP algorithm. The shortest path algorithm is not involved in this comparison because it only does route selection based on static parameters.
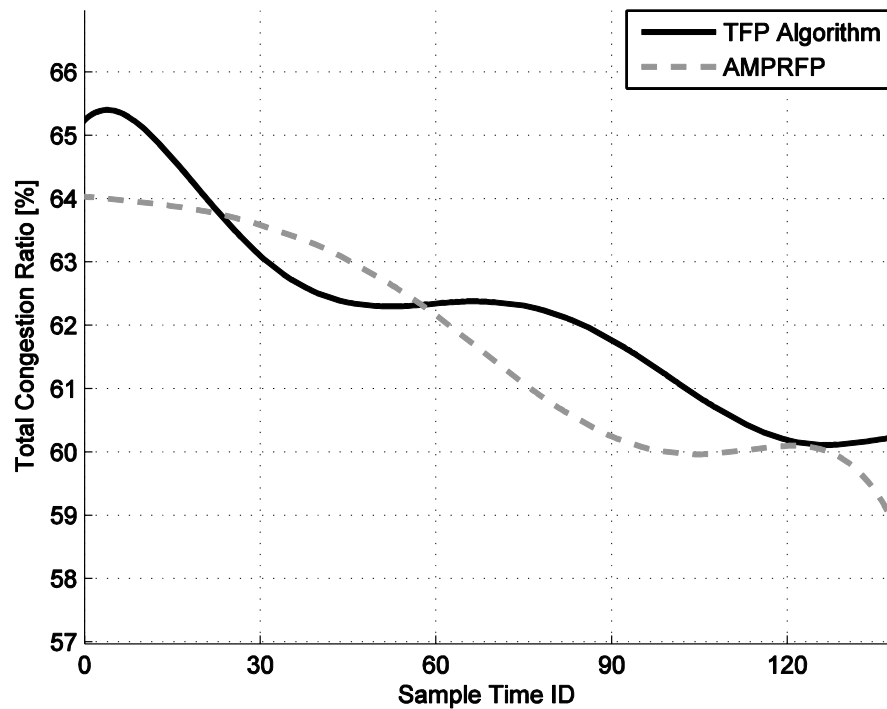

Figure 3.13: Comparison of route congested condition when using different route guidance algorithm.

In Fig. 3.13, the ratio of congested routes $(r = \frac{n_{congested}}{n_{total}} * 100\%)$ in the urban network during simulation using the AMPRFP algorithm and the TFP algorithm are

compared. With the same initial traffic condition, we observed that, at the beginning, the congested ratio of the TFP algorithm experiences a sharp increase. This is due to the TFP algorithm based system excluding congested arcs and guiding vehicles to the other arcs. In contrast, the system using AMPRFP algorithm does not roughly give up those congested routes but estimates the approximate travel time based on the analysis of the reason of each congestion. Thus, congestions in the whole network are solved slowly but constantly. Since the TFP algorithm based system will not send any vehicles to congested areas, after a period, the traffic density in those arcs reduced and the congestion ratio dropped too. During this period, the congestion ratio of TFP dropped faster than AMPRFP due to the release of traffic capacity in congested arcs. However, this phenomenon stopped when a new incident happened at sample time 47. The congestion ratio of TFP increased slightly since the number of vehicles in the urban network is less than in the beginning. While at the same time, the congestion ratio of AMPRFP continued to decrease along with the descent in the number of vehicles. During the entire simulation, only 20% of the time the system using the TFP algorithm had less congestion ratio compare to the AMPRFP algorithm based system. Additionally, the figure showed that the AMPRFP algorithm is able to distribute traffic flow more efficiently and more smoothly than the TFP algorithm. These results match our expectations of this system.

3.7 Summary

Real-time route guidance systems are required due to the continuing increase of demands of avoiding congestion, saving time, and saving money. As an extension of our previous work [68], in this paper, we have developed a congestion reason analysis based

77

real-time wireless route guidance system based on the AMPRFP algorithm to improve the traffic efficiency of a complex environment—a multi-lane urban area. This proposed system is developed to reduce the travel time of drivers' trip and to help them avoid congestion. It is also designed to overcome the inefficient diversion strategies problem by using real-time traffic data and the time prediction method.

The results of the field test show that our proposed AMPRFP algorithm, based on congestion reason analysis and iterative travel time prediction, can significantly increase the accuracy of predicted travel time during congestion situation by about 100%. This achievement solves the problem of current travel time prediction methods in that their performance drops sharply during congestion when they mainly rely on historical training data.

Additionally, the results of numerous simulation runs show that our system, which combined the historical traffic behaviors with the real-time non-linear effects caused by the neighboring arcs' traffic patterns, can achieve a dramatic trip time reduction. The performance of this system is more efficient than the one using the shortest path algorithm and the TFP algorithm by about 70.87% and 69.29% respectively. The results also show that the system can help to distribute traffic flow more rationally for the whole urban area and to improve the global optimization performance. Furthermore, it is able to release congestion more constantly than the system using TFP algorithm.

The disadvantage of this method is the time-lagged problem illustrated in Fig. 3.11. The predicted travel time based on AMPRFP algorithm will experience a delayed response to the change of the actual travel time in congestion involved scenarios. This is due to the time required for the analysis of reason of congestion after its occurrence.

Additionally, the highway patrol and emergency responders need extra time to verify the report of incidents using traffic cameras. This may be solved by employing new traffic monitor systems in the urban network that can provide nearly real time incident detection services.

CHAPTER 4

CELLULAR NETWORK LOCATION ESTIMATION VIA RSSI-BASED DATA CLEAN ENHANCED SCHEME

From chapter 3, we learnt that the AMPRFP algorithm based route guidance system can improve the traffic efficiency and help drivers avoiding congestion in urban traffic networks. One parameter which is essential to the success of the AMPRFP algorithm is the real-time traffic information. However, in the real-world, there is no efficient method to collect this information especially in local arterials.

In this chapter, we propose a data clean enhanced scheme to improve the accuracy of cellular network localization which is the important foundation of Floating Car Data (FCD) [92].

4.1 Motivation

Cellular network based context-aware applications and location based services (LBSs) have drawn significant attention in both research and industry for many years. A key aspect that influences the quality of context-aware applications and LBSs is the localization accuracy of the mobile terminal (MT). The empirical location estimation method, also known as the fingerprint method, is a popular location estimation technique proposed for providing high accuracy position results. According to this method, the observed signal is compared with signals of known locations in the database. The closest location is determined to be the position of the unknown target. Most of the proposed algorithms for this method focus on predicting coordinate vectors of an unknown location using various prediction models and the empirical data. However, the efficiency of this

method is questionable and no paper shows efforts on comparing the coordinates-based prediction method with the distance-based prediction method. In this paper, we present a data clean scheme enhanced empirical learning algorithm (DCSEEL) for improving the accuracy of the cellular network-based localization technique. It first minimizes the error that exists in the estimated distance between the target point and each reference base station (BTS). Then the trilateration method [10] is used to calculate joints according to these distances. Unqualified joints will be excluded through a direction filter (DF). The proposed algorithm is proved to outperform the coordinates based fingerprint techniques and is able to provide more accurate position results.

4.2 Background

4.2.1 Location Estimation Technologies

Location estimation is an emerging application in which noisy position-related parameters are measured from multiple channels to derive the location of the MT. The primary motivation of location estimation comes from the Federal Communications Commission (FCC) mandated requirement to provide location information for emergency services via Wireless Enhanced 911 in order to protect the health and safety of people and property. Additionally, the continuing increase of the demand of commercial LBSs and context-aware applications such as museum tour-guides, hospital health-care, and advertisement services, also play a key role in stimulating the development of a more accurate localization technique.

The global positioning system (GPS) had been recognized as one of the most successful localization techniques for many years [93]. However, its widespread use has

been limited by its reliance on special hardware which is not available in most mobile units. Furthermore, in metropolitan areas, the access to GPS signals is often constrained by the lack of direct line of sight (LOS) to satellites. Other handset-based solutions, such as Wi-Fi-based localization [94] and augmented sensor-based localization [95], have the same problem in that they require installing specialized hardware components to support cellular network localization services which are not widely applied in many phones. Thus, research for cellular network-based location estimation, which aims to be executable on all cellular phones, has drawn sustained attention in recent years.

The primary challenging aspect of network-based location estimation in cellular networks is to impute the channel noises which attenuate the field strength and spread the obstructed path of the transmitted waves. The source of that noise includes absorption, scattering, reflection, refraction, diffraction, multipath, and Non-Line-of-Sight (NLOS), which are defined as the Environmental-dependent Factors (denoted as EFs) [2]. Various research groups addressed this challenge and proposed different solutions to improve the accuracy of estimated location results [10]. These algorithms can be classified into two categories: a) signal propagation model approach [15, 96] and b) fingerprint approach [2, 4, 56, 59, 97]. All approaches try to minimize errors that exist in position related parameters caused by EFs. However, to build a universal propagation model for various environments relying on physical parameters of the received signal is quite difficult due to the complexities and diversity of environmental factors. This disadvantage has led more researchers to concentrate their efforts on developing fingerprint approaches. Fingerprint methods avoid the design of complex propagation model and are able to provide very accurate position estimation results even in a challenging environment. This

is achieved by using the empirical data observed in the training phase, compared to received position-related parameters at an unknown position. These parameters include radio signal strength (RSS), angle of arrival (AOA), time of arrival (TOA), and time difference of arrival (TDOA). Although time-involved parameters (TOA and TDOA) are more reliable than the signal-involved parameter (RSS), the installation requirement of location measurement units (LMU) at each BTS to provide precise time synchronization constrains its application. Therefore, RSS is selected as the qualified parameter due to characteristics such as common availability, economical efficiency, and compatibility.

4.2.2 Coordinates-based Approaches

Before introducing the detail of proposed DCSEEL algorithm, we briefly present the background of current coordinates-based approaches for cellular network location estimation. Three representative algorithms, namely Cell-ID, K Nearest Neighbors (KNN) deterministic fingerprint method, and Support Vector Regression (SVR) model-based fingerprint method, are used for comparison with the proposed algorithm.

A. Cell-ID

Cell-ID positioning concerns the location of the associated BTS as the location of the MT. In the cellular network, most phones constantly measure the signal strength from surrounded BTS's and record the closest 8 BTS's. The BTS which provides the strongest signal strength will be selected as the associated BTS to support general phone services. This method is very simple and does not need extra equipment in both the cell tower and the handset. It even works well in phones that cannot record closest reference BTS's but only the associated BTS. However, the accuracy of this method varied along with the

change of the cell size. Additionally, since the majority of cellular network cells are projected from an antenna with a spread of 120 degrees, the signal coverage area of a BTS may focus at a corner but not the center. Currently, Google's MyLocation and Ericsson Labs' mobile location API are using Cell-ID method to provide location services for mobile subscribers.

B. KNN

Generally, the process of KNN deterministic fingerprint technique can be divided into two phases: offline training phase and online tracking phase [56]. During the training phase, the measured signal strength of each BTS at different sample points with known coordinates is stored in pairs in the database. This is finished by selected subscribers whom periodically report the scanned cell tower ID, RSS, and their current GPS locations.

The KNN classification algorithm is executed in the tracking phase to classify training samples based on the closest fingerprint signature in the feature space. In order to implement it, the new signal reading is used to search the fingerprint database in order to find k closest samples with known locations. Then the coordinates of the current target are assigned as the weighted average position of its k nearest neighbors. This method can provide better localization result compare to Cell-ID in lower density areas, however, a large database is required to store training data set and a long offline training period is necessary.

C. SVR

Model-based fingerprint approaches focus on developing a model to define the relation function f(S) between the signal strength and spatial coordinates which satisfy

$Y = f(\mathbf{S}) + \varepsilon$. The set $\mathbf{S} = \{s_{\mathrm{q}i}\}$ is the whole signal strength vector where s_{qi} represents the RSS signature of BTS q at sample point i. The set $Y = \{x_i \parallel y_i\}$ contains one of the spatial coordinates x_i or y_i of the i_{th} sample. The variable set $\varepsilon = \{\epsilon_{qi}\}$ is defined as the offset scalar and satisfies $\varepsilon \subset \mathbf{R}$.

Many proposed algorithms [2, 10, 56] have contributed to developing model-based fingerprint techniques to solve cellular network localization problems. Along with them, the SVR algorithm shows advanced performance by utilizing kernel-based learning methods and the structural risk minimization (SRM) principle [2]. The key concept of SVR is its general belief that data can be more divisible when transferred to a high-dimensional feature space [62] and the nonlinear relationship between the input and the output can be more significant. The other important feature of the SVR approach is that the regression process only applied on the nonlinear expansion of the original input vector. This feature helps to maintain the property of the input vector. Compared with conventional empirical methods, the goal of SVR is to improve the generalized performance of regression by minimizing the upper bound of the expected risk but not the observed training error.

During the training phase, for a BTS q involved in the interested area, we assume there is an input vector $\mathbf{S}_q \subset \mathbf{R}^n$ stored in the database, where $\mathbf{S}_q = \{s_{qi}\}_{i=1}^{n}$, and n is the number of training samples collected in previous war driving. In order to build a generalized model relying on the dataset $\mathbf{S}_q$, a 1-norm loss function l, which is insensitive to slight deviations, is applied. Given a mapping function ϕ, the input space $\mathbf{R}^n$ is transferred to the high-dimensional feature space $\mathbf{F}$. We build a linear function f, to model

every transferred data $\phi(s_{qi})$ to one of its associated coordinates (we use y_i as an example, x_i takes the same procedure). The linear function can be written as:

$$y_i = f\left(\phi(s_{qi})\right) = \mathbf{w_q}\,\phi(s_{qi}) + b. \tag{4-1}$$

Then the next step is to decide the vector $\mathbf{w_q}$ and the constant variable b during the learning process. First we define the ϵ-intensive loss function l as:

$$l = \begin{cases} 0, & \text{if } \left|y_i - f\left(\phi(s_{qi})\right)\right| - \epsilon < 0 \\ \left|y_i - f\left(\phi(s_{qi})\right)\right| - \epsilon, & \text{otherwise} \end{cases}. \tag{4-2}$$

To simplify (4-1), slack variables ξ_i and $\hat{\xi}_i$ are introduced to cover otherwise infeasible limitations of the convex optimization problem. Thus the loss function can be written as:

$$l = \xi_i + \hat{\xi}_i \tag{4-3}$$

Since the linear function has the following characteristics: 1) for all training samples $\phi(\mathbf{S_q})$, f has no more than ϵ deviation exists with the matched known location vector $\mathbf{Y}$; 2) f needs to be as flat as possible, which means the Euclidean norm of $\mathbf{w_q}$ needs to be minimized. These characteristics can be implemented by solving following optimization problems:

$$minimize \; \frac{1}{2}\|\mathbf{w_q}\|^2 + C\sum_{i=1}^{n}(\xi_i + \hat{\xi}_i) \tag{4-4}$$

$$\begin{cases} y_i - \mathbf{w_q}\phi(s_{qi}) - b \le \epsilon + \xi_i, \\ \mathbf{w_q}\phi(s_{qi}) + b - y_i \le \epsilon + \hat{\xi}_i, \end{cases} \tag{4-5}$$

$$\xi_i, \hat{\xi}_i, \epsilon, C \ge 0, i = 1,2,\dots,n. \tag{4-6}$$

In (4-4), the parameter C is used to balance the flatness of f and its ability to tolerate deviations. Its value is given before the training phase. After maximizing the corresponding dual problem, the vector $\mathbf{w_q}$ can be represented by following expansion:

$$\mathbf{w_q} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\,\phi(s_{qi})$$
(4-7)

Then the linear function can be

$$\begin{cases} y_i = f\left(\phi(s_{qi})\right) = \sum_{j=1}^{n}(\alpha_j - \alpha_j^*)\,k(s_{qi}, s_{qj}) + b \\ k(s_{qi}, s_{qj}) = \phi(s_{qj})\phi(s_{qi}) \\ \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C] \end{cases}$$
(4-8)

where k is the kernel function that determined the quality of mapping.

In the tracking phase, new signal readings will be mapped into the feature space and the trained linear function will be used to predict the possible value of corresponding coordinates.

4.3   DCSEEL Algorithm

4.3.1   Motivation

One of the critical problems of current empirical learning methods is that the geometry feature (latitude, longitude, etc.) of position-related parameters is not well considered. In a geometrical map, two main parameters, distance and angle, comprise the position of a unique point. Therefore, in the cellular network, signal strength defines the irregularly shaped area that is centered by the referenced BTS. However, using the coordinates-based fingerprint method, a particular target point surrounded the selected BTS is determined only based on the received signal strength. Unless the density of training samples are extremely high and can totally cover all places in the testing area,

these methods are facing potential risks in locating a wrong position due to the incompleteness of the training dataset.
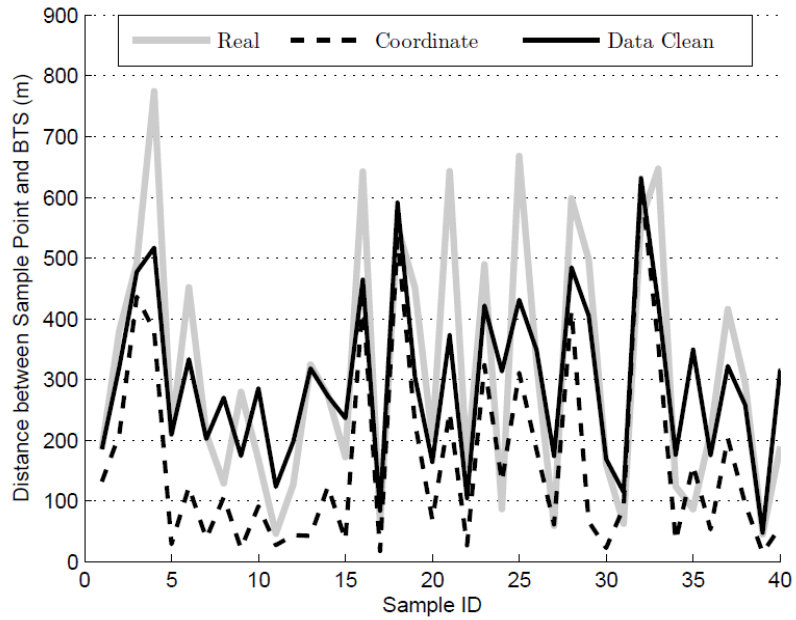
(a) Distance Comparison
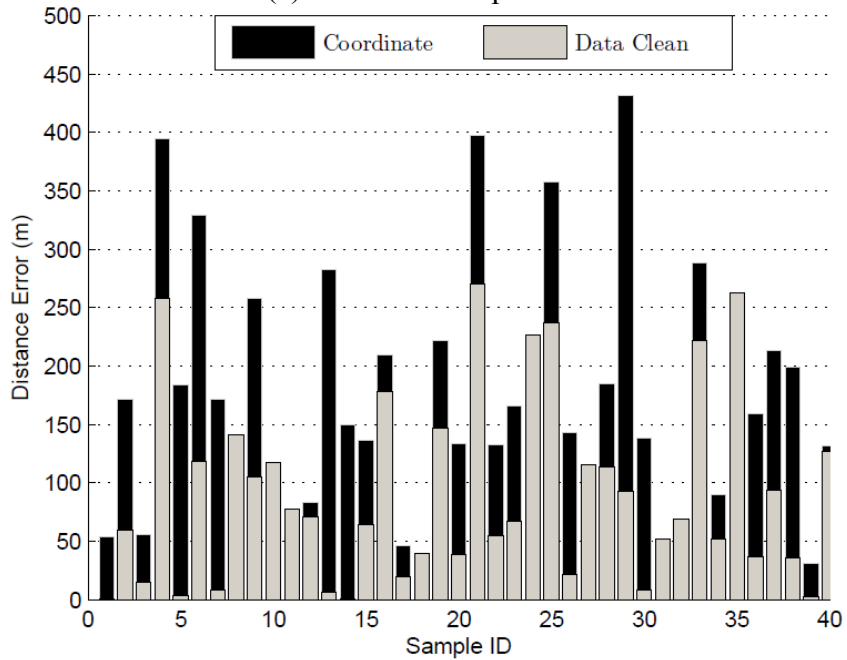
(b) Distance Error Comparison

Figure 4.1: Comparison of distance from sample point to BTS by different methods: (a), (b).

Unlike the coordinates-based fingerprint method, the data clean enhanced method tries to minimize the errors by relying on received signal strength and historical data. In this way, more accurate distances will be predicted and joints of these less error spheres will narrow down the error range of the estimated final position.

Fig.4.1 compares the distance calculated by the predicted coordinates by SVR and the target coordinates and the distance predicted using SVR directly. The detail of the synthetic experiment process will be introduced in the following section. From Fig.1.1.(a), we can figure out that the distance results predicted by SVR directly are closer to the real distance between the sample point and the BTS than the distance calculated by the predicted coordinates. In Fig.1.1.(b), the distance error of these two methods also shows the same phenomenon, since the black bar which represents the distance error of coordinates based method has more opportunities override the distance error of the data clean enhanced method.

Table 4.1: Performance Comparison of Predicted Distance Error

Round ID	MAE_COOR [m]	MAE_DC [m]	Ratio [%]
1	123.3314352	88.32736274	65.75
2	126.9837609	85.32129866	69.00
3	128.1812705	85.06334425	66.75

Table 4.1 shows the ratio of this comparison for 3 rounds, each round took 50 samples, each sample compared the distance error with its 8 reference BTS's. In this table, the MAE_COOR represents the mean absolute error of the coordinates-based method while MAE_DC means the mean absolute error of the data clean enhanced method. The result shows that in over 65% of instances in each round, the distance error decided by coordinates based method is bigger than the data clean enhanced method. This result

gives us confidence that data clean enhanced method is superior and is possible to provide better localization results.

4.3.2 Optimal Performance

In order to further verify the inference that the data clean enhanced fingerprint method can perform better than the coordinates-based fingerprint method, we try to compare the optimal result achieved by these two methods. The optimal result means to find out the combination of points or joints which average coordinates has the least distance to the real position of the test sample. An optimal filtering algorithm, modified from Dijkstra's algorithm [85], was developed to determine the optimal result.


Algorithm 1 OPTIMAL FILTERING(V, E)

1: **for** each vertex v in **G do**
2:      $d[v] := \sqrt{(lat[v] - lat[s])^2 + (lon[v] - lon[s])^2}$
3:      alat[v] := average latitude of **J** exclude v
4:      alon[v] := average longitude of **J** exclude v
5:      $cost[v] := \sqrt{(alat[v] - lat[s])^2 + (alon[v] - lon[s])^2}$,
6:      $r[v] := v$
7: **end for**
8: **J** := the set of all joints
9: **while J** is not empty: **do**
10:      j := joint in **J** with smallest $d[j]$
11:      remove j from **J**
12:      **for** each neighbor v of j: **do**
13:          alat[v]:= average latitude exclude j
14:          alon[v]:= average longitude exclude j
15:          tmp := distance from new average point to target
16:          **if** $tmp < cost[v]$: **then**
17:              cost[v] := tmp
18:              r[v] := j+r[v]
19:          **end if**
20:      **end for**
21: **end while**
22: j := joint in **J** with smallest cost[j]
23: return cost[j], j, r[j]

In the first step, we model the map of all joints of a sample point s as an undirected graph G(**V**,**E**), where the set **V** of the graph's vertices indicates joints and an edge between vertex i and j represents the distance between s and the average point excludes joint i and j and vice versa. The operation of the optimal filtering algorithm is summarized in the following pseudocode. *lat*[i] and *lon*[i] represent the latitude and longitude of joint i while *alat*[i] and *alon*[i] represent the average latitude and longitude of the group of joints without joint i respectively. *r*[i] records the excluded joints list that contribute to the reduction of distance error *cost*[i].

(a) Urban

(b) Suburban

91
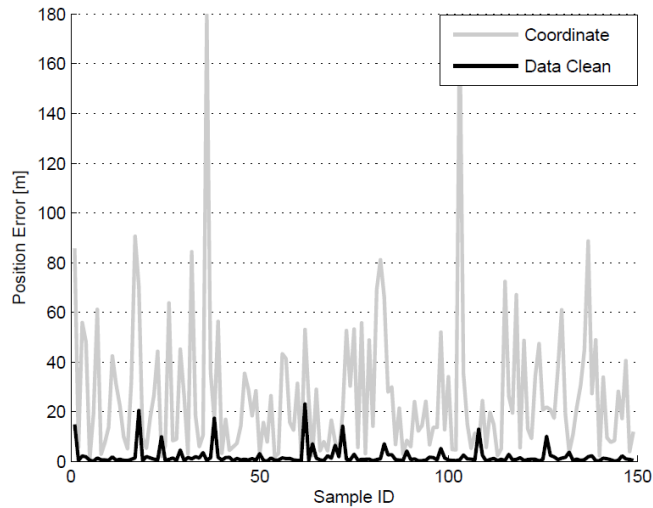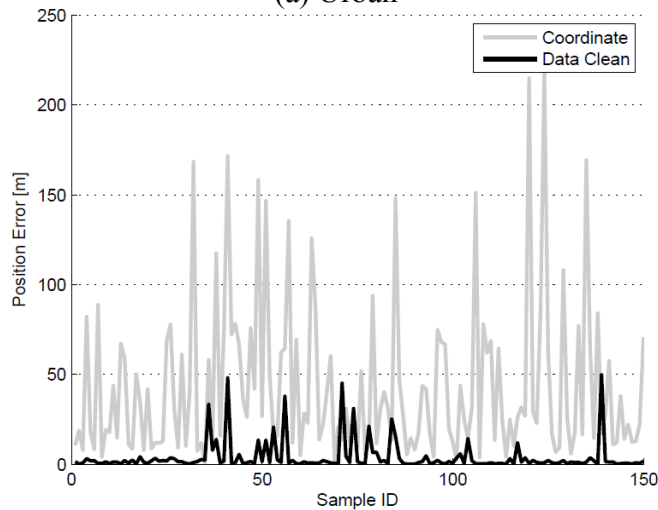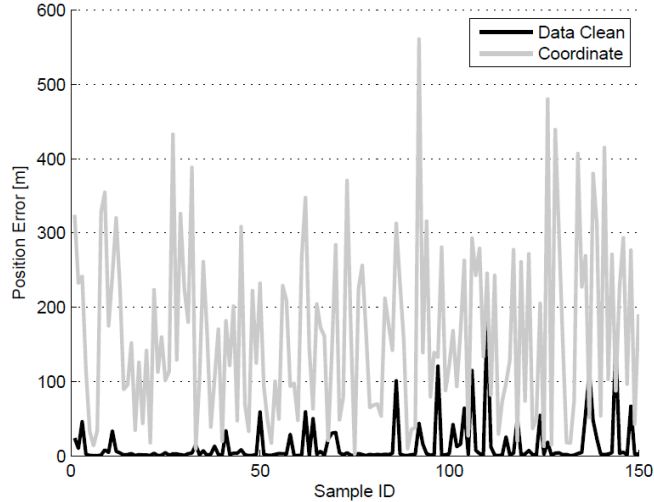
(c) Rural

Figure 4.2: Comparison of optimal localization results: (a), (b), (c).

We take 150 samples for each environment: urban, suburban, and rural. In Fig.4.2, the gray line represents the position error of the estimated optimal location of each sample applying coordinates-based SVR, while the black line shows the position error of our data clean enhanced SVR. From this figure, it is obvious that the position error of the coordinates based SVR is more variable and fallible than the data clean enhanced SVR. In the three environments, more than 95% of the optimal results of data clean enhanced SVR are less than coordinates-based SVR. The detail data comparison is shown in Table 4.2.

Table 4.2: Predicted Position Error in Optimal Situation

Enviroment	MAE_COOR [m]	MAE_DC [m]	Ratio [%]
Urban	26.2981555	1.920268944	97.33
Suburban	43.03683765	3.890516306	94.67
Rural	164.3599499	13.54529953	96.67

4.3.3   Initial Direction Decision

In the previous section, we show the encouraging optimal localization results that can be achieved with the knowledge of the target point. However, in the real world, the

position of a sample point is unknown. Therefore, determining an efficient parameter in lieu of the target position is critical to the performance of our DCSEEL. Since the final goal of this method is to clean out inappropriate joints from the whole joints group, in the geometric phase, the most important problem is to find a method to lead the average point close to the target point following a specific path. In other words, if we can find the direction from the original average point calculated by the whole group of joints to the target point, then the joints that need to be cleaned are those that will make a contribution in maintaining the average point following this initial direction to approach the sample point. As shown in Fig.4.3, for example, the direction from *Avg1* to target which is marked with the dashed line, is set to be the initial direction. *Avg2*, *Avg3*, and *Avg4* are average points follow initial direction and will get closer to the target, while *Avg5* which go against the initial direction has a bigger location error compare with *Avg4*.
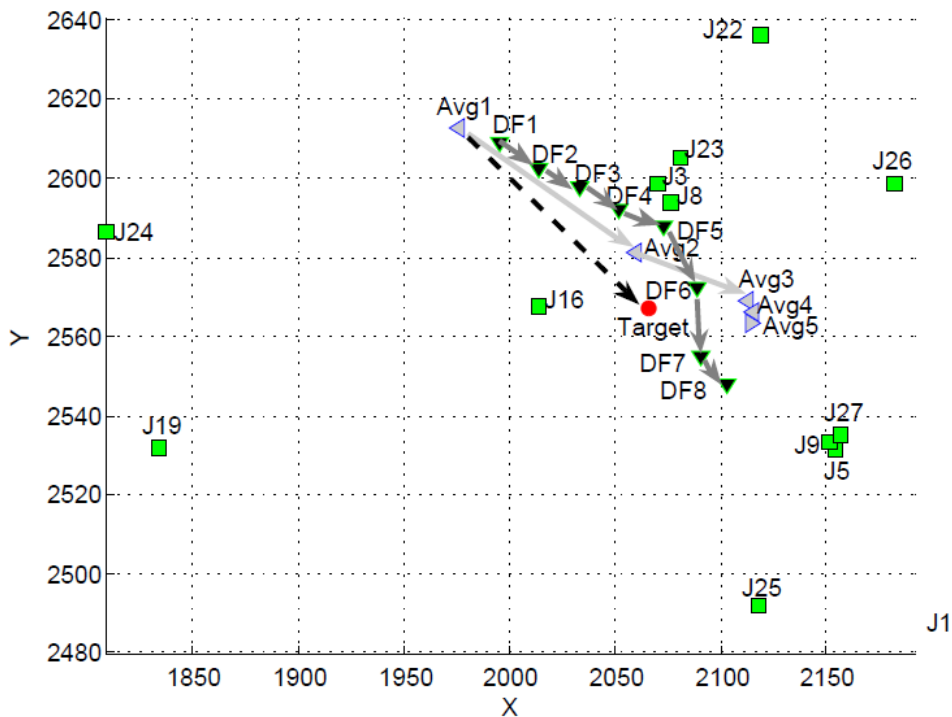

Figure 4.3: Initial Direction.

In order to determine the initial direction, a cluster method is applied. Every time we decide the average point of the selected group of joints **J**. After this, we calculate the distance between each joint to the average point and the average distance among these distances. All joints in **J** with a distance larger than the average distance will be excluded. The whole process is repeated until the new average point change direction. Finally, the initial direction is determined.

4.3.4   Direction Filtering (DF)

The initial direction decision process is able to decide the direction from the original average point to the target. However, this method is so crude that it cannot achieve high quality results. Therefore, a DF method is used to construct the remaining set **R** by excluding all joints that will contribute to the movement of the average point in **R** close to the target. This is achieved by combining the moving direction and the moving distance together to decide which joints should be excluded and when to terminate this process. At the beginning, where **R** = **J**, we calculate the average coordinates for each joint $j \in \mathbf{R}$ by excluding j from **R**. Then, we want to find

$$\max [D(j)] \tag{4-9}$$

where *D(j)* is the distance between the average point excluding j to the previous average point j^{*}, following the initial direction. This method has benefits such as helping the average point to avoid frequent fluctuation and reducing complexity of the design of this filtering method.

An auxiliary parameter, *mainD,* defined as the sub-direction with the max offset in horizontal or vertical orientation. This parameter is important in overcoming the

94

over-fitting problem and in deciding the time to terminate the whole filtering process. If the *mainD* maintains the same value until no average point of joint $j \in \mathbf{R}$ falls onto the initial direction, the DF process is terminated and the average point decided by **R** is the final estimated position. Otherwise, if the *mainD* changed more than once, the over-fitting strategy is triggered and only joints excluded in the first two *mainD* phases are removed and other joints will be kept. To illustrate this, let us consider the same example depicted in Fig.4.3. The movement of the average point from *Avg1* to *Avg4* is following the initial direction. However, from the figure we can learn that *Avg3* already goes across the target point in the initial direction and the farther the average point moves, the bigger distance error will be achieved. Therefore, the result of the DF method, represented by black triangles, illustrates a smoother movement and can achieve less distance error.

4.4 Synthetic Experiment

In this section, we study the ability of DCSEEL in reducing localization errors and compare its performance to other localization algorithms.

4.4.1 Set Up and Methodology

To assess the performance of DCSEEL, we select some RSSI-based cellular location estimation methods, such as the Cell-ID, 3NN, and SVR, as reference algorithms. The synthetic data generator is designed by using the generative propagation model STA-EM which was proposed in [96]. The simulation area for signal acquisition is set to be a square region with a size of $1,000m^2$. From [2], we learned that the MT in the testing area may receive signals from BTS's located in the surrounding area which is larger than the

testing field. Thus, we deployed the reference BTS's in a 5,000m^2 area and set the testing field to be the center of this area. In order to test the performance of our proposed algorithm in different situations, three test neighborhoods, an urban area, a suburban area, and a rural area were created. Based on the field test data of [98] in Seattle, the city with the densest distribution of cellular towers in the US, we set the average cell density of the urban environment as 66$cells/Km^2$ and the suburban area as 26$cells/Km^2$. For the rural area which is not tested in [98], the average cell density is set as 2$cells/Km^2$ which is similar to the records of the Everglades National Park area in Florida. These setting include nearly all situations of the outdoor cellular network from the densest area to the sparsest area.

For each testbed, 3 instances of simulation were executed. And for each round, there were 2000 samples generated in advance for training purposes. Then 50 samples were created as the test samples. The position of each sample was selected randomly from the testing field and 8 signal readings from reference BTS's were calculated by the STA-EM model. In this research, the problem of missing values is not considered and we assume that at every time, the sample MT can always hear 8 signals from surrounded BTS's.

Parameters in SVR are specified in the following manner: The kernel function used in SVR is selected as the analysis of variance (ANOVA) due to its advanced performance over other kernel functions such as dot, polynomial, and radial. Parameters of ANOVA, d and γ, are specified by the test in [68] in order to achieve optimal testing results. Other parameters were selected relying on empirically favorable values: $C = 1,000$; $\epsilon = 0.01$. The implementation of ANOVA based SVR is using a mature software for SVM research named mySVM [87].

4.4.2   Ability in Reducing Localization Error

As mentioned in section 4.3.2, optimal results of the data clean enhanced method show great improvement in reducing localization error compared with the coordinates-based method. However, achieving optimal results is impossible due to the lack of knowledge of target. In this section, we explore the suboptimal result of our proposed DCSEEL and the traditional coordinates based method.
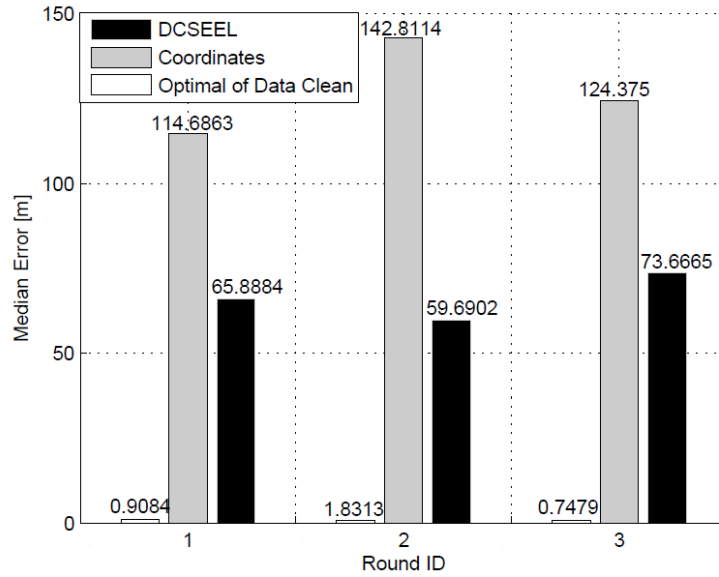
(a) Urban
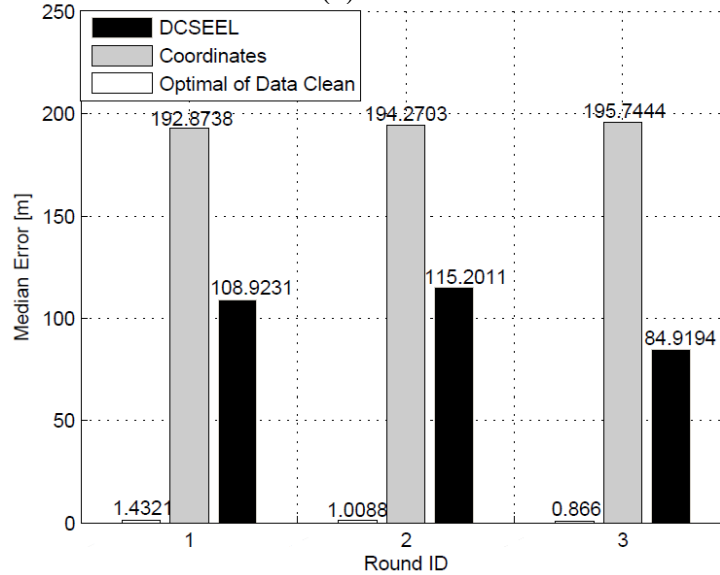
(b) Suburban

(c) Rural

Figure 4.4: Performance of data clean enhanced method and coordinates based

fingerprint method: (a), (b), (c).

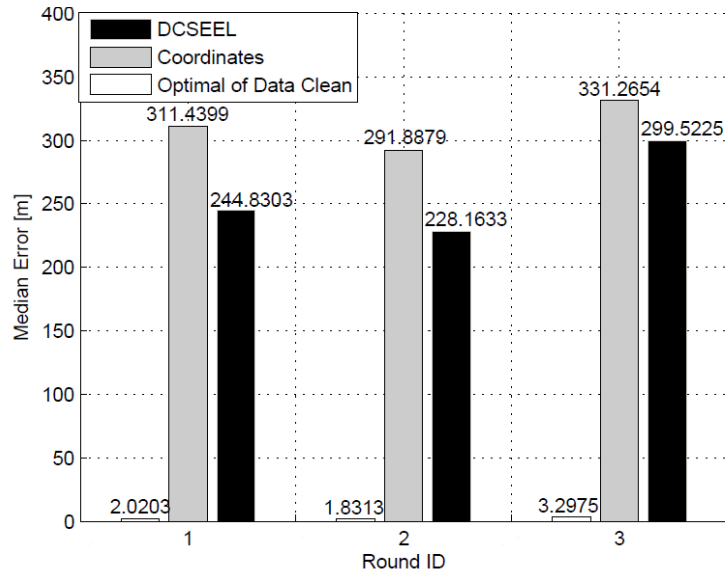Figs.4.4 (a),(b),(c) show the median localization error for each round in different testbeds: urban, suburban, and rural. These metrics provide evidence that in urban, suburban, and rural environments, the DSCEEL can provide more accurate localization results and the rate of improvement is no less than 40% in both urban (round1: 42.55%; round2: 58.2%; round3: 40.77%;) and suburban areas (round1: 43.53%; round2: 40.7%; round3: 56.62%;) and no less than 10% in rural areas (round1: 21.39%; round2: 21.83%; round3: 10%;). These results prove the previous assessment that the data clean enhanced fingerprint method is better than the coordinates-based method. The improvement dropped in the worst case of the rural area due to lower density of cell towers which increases the difficulties in cleaning errors that exist in received signal strength.

4.4.3   Comparison with Other Techniques

In this section, we compare the overall performance of DCSEEL, in terms of cumulative distribution function (CDF) of localization error, to other representative algorithms described in Section 4.2.2.
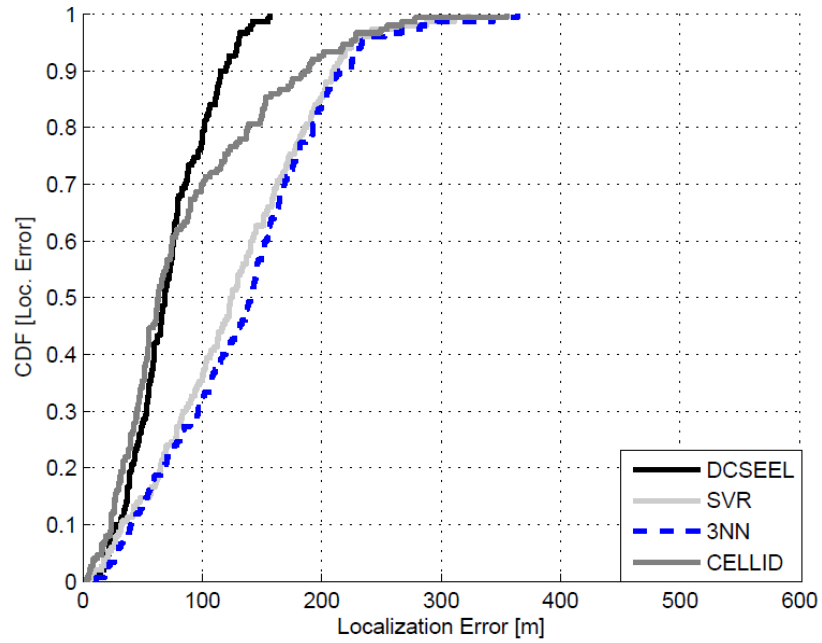
(a) Urbna
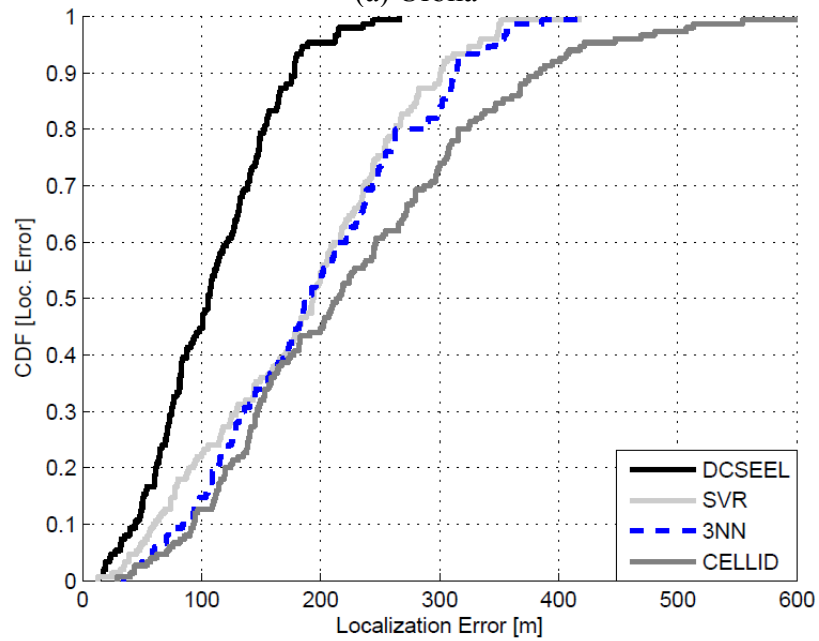
(b) Suburban
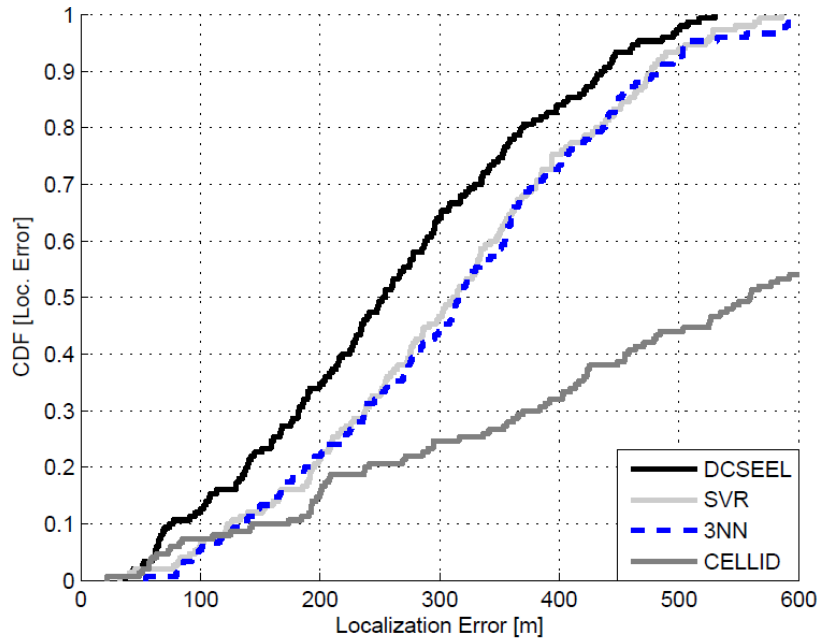
(c) Rural

Figure 4.5: CDF's of distance error for different techniques under different

environments: (a), (b), (c).

Shown in Fig.4.5, the CDF of localization error estimated by different algorithms in different environments are compared. It can be seen from Fig.4.5(a) that in the urban environment, where the density of cellular cells is very high, the Cell-ID method has similar performance at the probability space [0, 0.6] to DCSEEL over other localization methods. From the probability space [0.6, 1], DCSEEL performs better than cell-ID method. For two RSSI-based fingerprint methods, SVR shows better performance than the 3NN method due to its ability in minimizing the upper bound of the expected risk and improving the generalized performance. It is distinct from Figs.4.5 (b) and (c) that in both suburban and rural environments, the DCSEEL provides the best performance among the four compared algorithms. The performance of the Cell-ID method dropped sharply when the density of cell towers decreased remarkably. Two fingerprint methods, SVR and 3NN,

maintain a constant performance in these two environments, since the accuracy of these two methods highly depends on the quality of the training dataset.

4.5 Summary

In this paper, the problem of location estimation in outdoor cellular networks has been addressed. A data clean scheme enhanced RSSI-based empirical learning algorithm for cellular mobile terminals was developed and tested. We presented the details of the scheme and how it constructs the remaining set based on joints of predicted distance spheres. Our results show that DCSEEL can provide accuracy improvement with an average value of 47.17% in urban areas, 46.95% in suburban areas, and 17.74% in rural areas. We also studied the general performance of DCSEEL with other techniques by comparing the CDF of them. The DCSEEL also shows advanced performance in both environments.

CHAPTER 5

FAST NORMALIZED CROSS-CORRELATION ENHANCED FLOATING CAR DATA ESTIMATION

In this chapter, we propose a novel Floating Car Data (FCD) estimation method to improve the efficiency and accuracy of the cellular probe based speed estimation technique. A fast matching algorithm which focuses on reducing the computational cost of the matching phase in the correlation algorithm based cellular probe speed estimation [7] is investigated.

5.1 Motivation

This paper introduces a fast matching algorithm to reduce the computational cost of the matching phase in the correlation algorithm based cellular probe speed estimation [7].

As mentioned in Chapter.3, traffic congestion avoidance in urban areas becomes critical due to the fast development of cities and the persistent increase of the number of motor vehicles all over the world. The annual urban mobility report [67] showed that the waste of gasoline in the United States is more than 10 billion gallons every year. The total cost of economic inefficiency at the same period exceeds 63 billion dollars due to traffic congestion. To relieve congestion, one simple method is to build more routes and to widen existing ones. However, this method is quite expensive due to the requirement of a long construction period and a huge budget. Therefore, the most effective method to solve the congestion problem is to build intelligent transportation systems which can manage the traffic smartly. In our previous work [68], we concluded that with the support of real-time traffic information, the intelligent transportation system is able to efficiently

reduce the travel time of drivers and suppress the occurrence of traffic congestion. Additionally, real-time traffic speed is essential in current intelligent transportation systems for identifying traffic congestion and providing high quality navigation services. Therefore, the development of an efficient real-time traffic information collection method is urgent and important.

The correlation algorithm shows superior performance over the localization algorithm and the handoff algorithm in both highways and local arterial. According to this method, the recorded handset's signal strength profiles are compared with training traces at the same road. The speed scale, which is determined by the stretch or compression rate of the matched training trace, is used to identify the speed of the target mobile probe. However, a critical issue of the current correlation algorithm, the time efficiency, was not investigated and discussed. We addressed these challenges by converting the matching process into a 1-dimensional (1-D) curve matching problem [99] and introducing the fast normalized cross-correlation (FNCC) algorithm [8] to reduce the computational cost. For a matching duration of size N and a training trace of size M, the time complexity of matching phase can be reduced by the Fast Normalized Cross-Correlation (FNCC) algorithm from $3N(M - N)$ additions/subtractions and $2N(M - N)$ multiplications to $9Mlog(M)$ additions/subtractions and $6Mlog(M)$ multiplications by applying fast convolution algorithms and sum-table technology. Experiment results concluded that 97% computational cost of the Pearson product Moment Correlation Co-efficient (PMCC) algorithm based matching method can be saved by implementing the FNCC method. The proposed algorithm is proved to outperform the PMCC method [9] and is able to maintain the accuracy of the estimated speed at the same time.

5.2 Real-Time Traffic Data Collection Techniques

5.2.1 in-situ Technology

The most widely used methods for real-time traffic data collection is the "in-situ" technology [92], which focus on collecting traffic information through various kinds of detectors with fixed locations along the roadside. Based on the observation method, this technology can be divided into two groups: intrusive and non-intrusive. The sensor or data recorder are placed on or in the road for intrusive methods. Some popular intrusive approaches are the pneumatic road tube [92], the inductive loop sensors [100], and the magnetic loops [92].

In contrast, the non-intrusive methods collect road data using remote methods, including the manual counts [101], the passive and active infrared detectors [102], the passive magnetic detectors [103], the microwave (Doppler) radar [104], and the video image detection [105].

Due to the direct access of the monitored road, the "in-situ" technology is able to provide highly accurate traffic flow and speed data with less positioning error. However, the huge cost of the installation and the maintenance of detector systems limit the coverage of this system which is valid only in major freeways and highways. For instance, in Kansas City, the installation of an inductive loop sensor system for 70-mile highway cost about $15 million [92]. In order to cover more fractions of highways and local arterial at lower cost, the Floating Car Data (FCD) technology [92] was proposed.

5.2.2 FCD Techniques

Different from the "in-situ" technology, which uses fixed detectors, FCD concentrates its effort on collecting real-time information from moving objects over the entire traffic network. The most important superiorities of FCD are the low cost implementation and the easy deployment along the road network due to its characteristic that does not require additional hardware installed on the roadside. However, since there is a lacking of predefined position of each floating probe, the location error cannot be ignored and the system needs further support of advanced localization technologies. Recent studies [80] have shown that GPS can achieve localization accuracy within 10m and is able to provide lane level localization services. Although GPS-based techniques show advanced performance in reducing position deviation, some drawbacks of it, which will influence its implementation, cannot be ignored. First, the number of GPS-equipped devices is still too small to satisfy the requirement for monitoring an entire city. Second, the high power consumption characteristic of GPS-equipped devices also limits the implementation of it on providing persistent real-time traffic information. Additionally, the accuracy of GPS dropped sharply in some special places or weather conditions, such as the urban canyon, cloudy weather, etc.

In order to overcome limitations of GPS-based speed estimation methods, cellular probe based approaches become particularly attractive due to its seamless coverage in the city and its small usage of battery. Many companies and organizations, including CELLINT, DELCAN, APPLIED GENERICS, AIRSAGE, and GLOBIS DATA, built a number of test projects for justifying the performance of cell phone FCD systems in the real world [106]. Most of their systems are based on the localization algorithm [107] or

the handoff algorithm [108], which tried to record the position of sampled points and calculate the average speed using the distance between two samples divided by the sampling period as shown in Fig.5.1. Unfortunately, the localization error of current cellular network location estimation methods is still too big to provide high quality speed results.
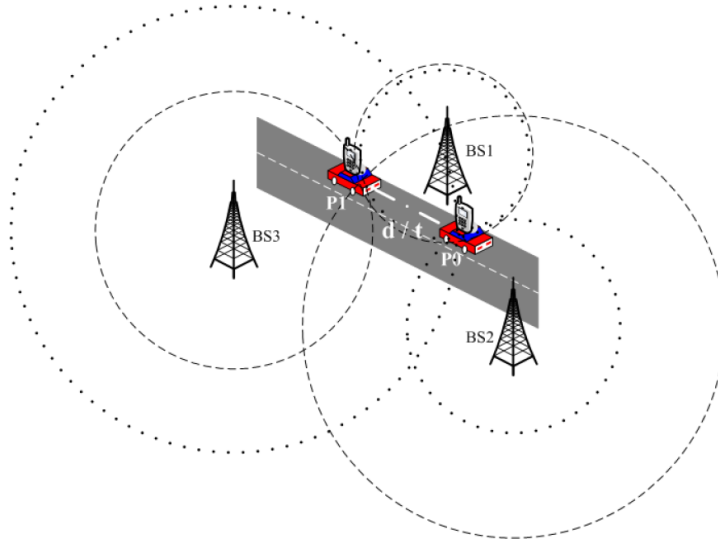
(a) Localization Algorithm
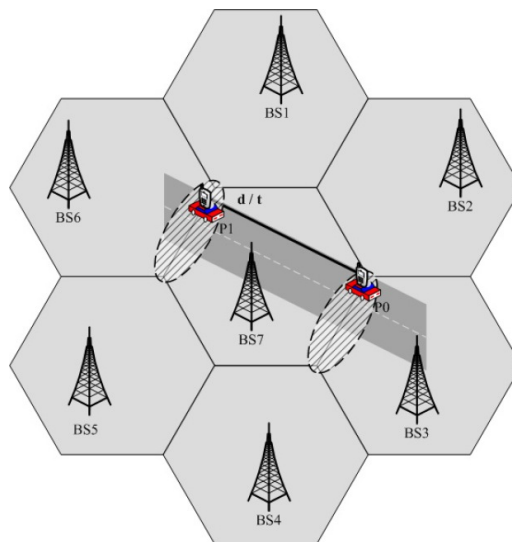
(b) Handoff Algorithm

Figure 5.1: Current cellular probe speed estimation algorithms: (a), (b).

5.3   Correlation Algorithm based Cellular Probe Speed Estimation

Different from the previous localization algorithm and the handoff algorithm shown in Figs.5.1(a),(b), Chandrasekaran et al. [7] first proposed a novel correlation algorithm based cellular probe speed estimation method, which determines speed by matching the test signal strength trace with the scaled training trace at a certain road and a specific point. The performance of the correlation algorithm based method is proved to be superior to the other two approaches in both highways and local arterial. It also works without the limitation of phone models and the network service carriers. In the following sections, we will give a deep introduction about this algorithm.

5.3.1   Signal Strength vs. Speed

In Chandrasekaran's paper [7], they discovered that the tracing of signal strength captured in a road with different speed showed an amount of stretching or compression in its scale.

As shown in Fig.5.2, a field test at the same road with constant speed was executed. For trace 1, the driving speed is about 25mph. The other two traces, trace 2 and trace 3, are both with speed of 50mph. It can be seen that the wave of two signal traces for the same speed are quite similar to each other. The signal wave of trace 1, which has speed of 25mph, shows a similar feature of the other two traces, except for being stretched by a scale factor of 2. This illustrates that the average signal strength at one position of a road remains relatively steady, even when there are signal variations. Similar phenomenon was also reported previously in [98] for stationary handsets and in Gundlegard's research [108].

Thus, in this step, the training trace of each road at speed *V* needs to be obtained in advance (this can be achieved when service carriers try to determine their coverage). Then, the *n* scaled version of the training trace can be generated. The number *n* is an integer which represents the maximum speed (we assign *n* = 80 in this research which represents 80mph) and the length of each step is set as 1mph. This scaled version depends on the accuracy of the system's demand.
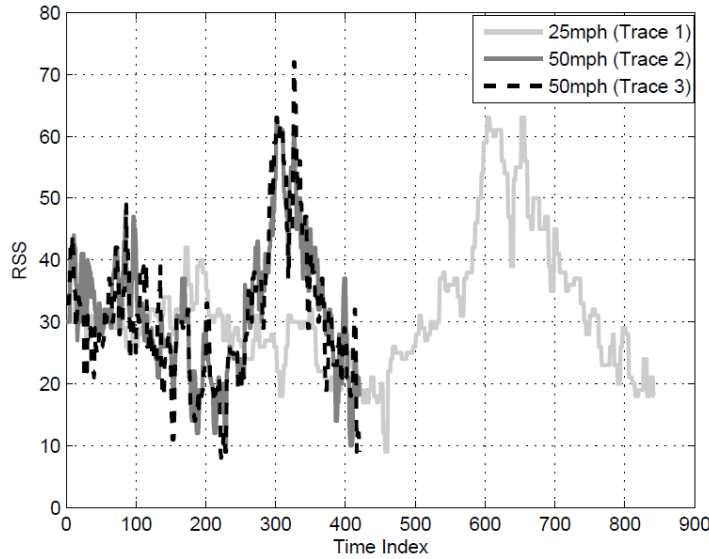
Figure 5.2: Relationship between the Signal Strength and the Speed.

5.3.2 Matching Phase

The core component of the correlation algorithm is the matching phase which focuses on comparing the input testing trace with the n scaled versions of the training trace at the same road and the same start point. In [7], the PMCC [109] was selected to estimate the matching level between each scaled training trace and the testing trace. The equation of the PMCC is shown in (5-1):

$$r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}}$$

(5-1)

where X_i and Y_i belong to two signal traces **X** and **Y** with the same length of *N*. The PMCC has high accuracy in determining the correlation of two objects. However, since it needs to calculate the correlation of all discrete nodes in the trace, the computational cost is very high that does not qualify for supporting real-time speed estimation services under the network based architecture.

Furthermore, this method requires the knowledge of the exact start point since the length of the two objects needs to be the same. This is quite difficult to be achieved in the cellular network using the network based location estimation algorithm. In [7], Chandrasekaran provided a method by defining a searching range S (shown in Fig.5.3) to overwhelm the error that exists in the localization result. For the urban area, the localization error is within 100*m* range centered with the estimated position. Thus, the value $S_j \in \mathbf{S}$ is an integer value calculated by (5-2)

$$S_j = \frac{100}{v_j} \times \frac{3600}{1609} \tag{5-2}$$

As shown in Fig.5.3, the start point is varied within the searching range **S**. And the PMCC value *r(i)* of each start point *i* is calculated independently. Unfortunately, this method is still not efficient in reducing the computational time of the matching phase.

5.4 Fast Matching Algorithm

5.4.1 Motivation

In the PMCC algorithm, the calculation of the coefficient for each point in the searching range **S** contains many redundancies. In order to make the process more efficient, we convert this problem into a whole-to-part matching problem. As illustrated

in Fig.5.3, it can be seen that the testing trace is potentially embedded in one scaled version of the training trace. Thus, we do not need to simply repeat the same matching process for each start point but to treat the training trace as a whole curve and do comparison with these two.
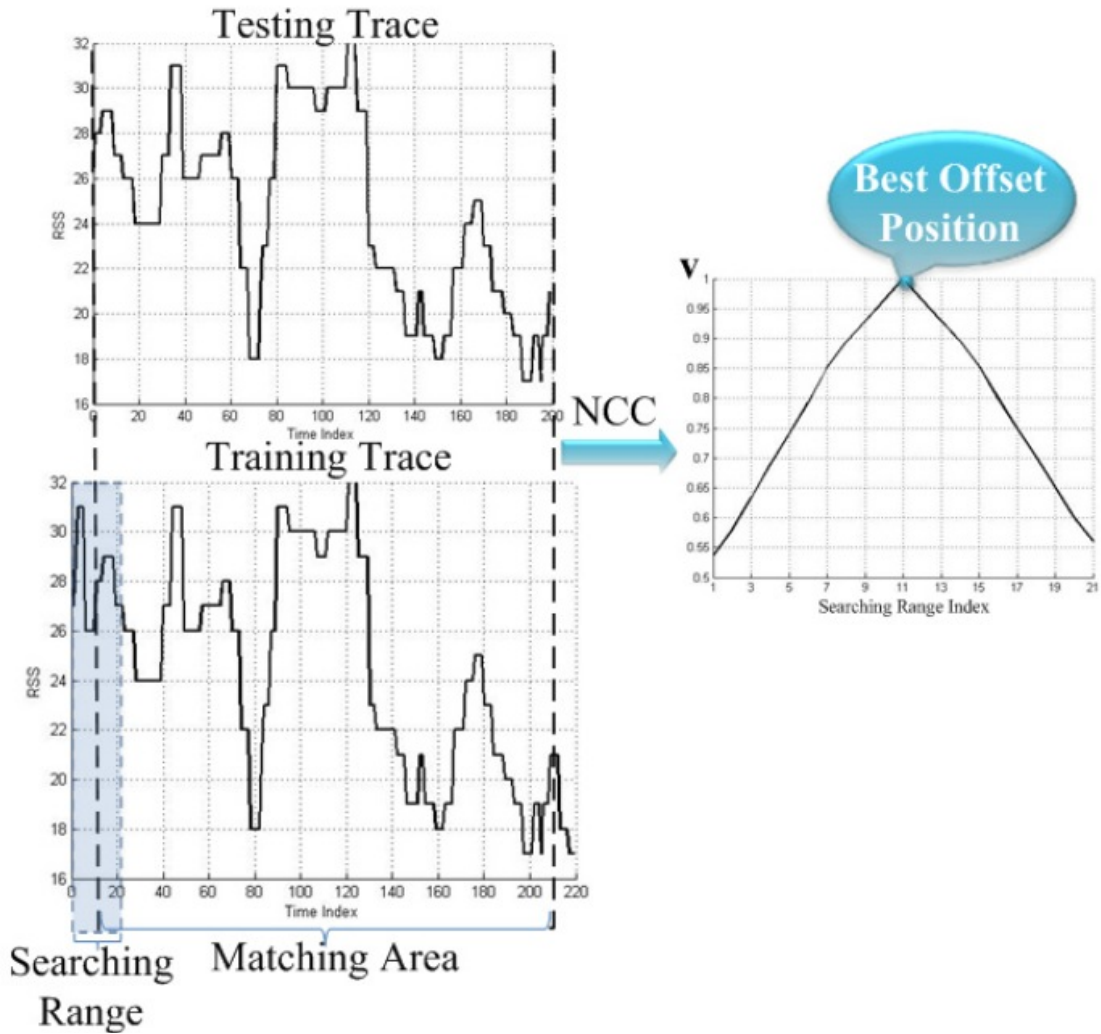

Figure 5.3: Example of 1-D whole-to-part curve matching.

5.4.2   Normalized Cross Correlation

Normalized cross correlation (NCC) [8] is famous in image-processing applications due to its ability in normalizing the brightness of the template image *t(i)* (size is *N*) and

the reference image *f(i)* (size is *M*). Different from the cross correlation (CC), NCC limited the output value *v(I)* lies in a standard area [-1, 1]. Additionally, the length of two objects, *N* and *M,* do not need to be the same. In most cases, the length of the reference image is larger than the template image, which waives the start point limitation of the PMCC. The formulation to calculate the NCC between two figures in each sample index $I\ (1 \leq I \leq M - N)$ is

$$v(I) = \frac{\sum_{i=1}^{N}[t(i)-\bar{t}][f(i-I)-\bar{f}]}{\sqrt{\sum_{i=1}^{N}[t(i)-\bar{t}]^2 \sum_{i=1}^{N}[f(i-I)-\bar{f}]^2}}.$$

(5-3)

This process can be seen as the template curve sliding along the reference one, as the same example shown in Fig.5.3. $\bar{t}$ and $\bar{f}$ are mean values of the template image and the reference image, respectively. As shown in the right part of Fig.5.3, the larger the value *v(I)*, the better the match is. Once we find the position *I* with the maximal output, we can decide the relative alignment of the two curves easily.

However, since NCC needs to check all sample indexes in the reference image, the computational cost is quite high that contains *2N(M - N)* multiplications and *3N(M - N)* additions/subtractions. Therefore, an efficient method to reduce the time complexity of NCC is necessary, especially when applying it to provide real-time services.

5.4.3   Fast Normalized Cross-Correlation

The problem of the NCC is its high computational cost. This is due to the implementation of the exhaustive search, which entail the template image contrasted with every possible comparison within a pre-specified searching range. In order to improve the efficiency, Lewis et al. [8] proposed a fast normalized cross-correlation method (FNCC)

by using pre-calculated sum-tables in the denominator and using the Fast Fourier Transform (FFT) in the numerator. This is based on the fact that most calculations are redundant due to the exhaustive search of the reference image and high overlap between the template images. In the following section, we will discuss the optimization of the denominator and the nominator, respectively.

A. Denominator

The denominator includes two sections: the standard deviation of the template image and the standard deviation of the compared image along the reference figure. For the template image, the value of the standard deviation in $\sqrt{\sum_{i=1}^{N}[t(i)-\bar{t}]^2}$, is a constant which can be calculated in time *3N* in advance. Therefore, the time complexity of the denominator can be reduced by optimizing the calculation process of the following expression: $\sqrt{\sum_{i=1}^{N}[f(i-I)-\bar{f}]^2}$ at each sample index *I*.

Assume we have $F(i) = f(i) - \bar{f}$ and the *M* value of *F* can be calculated in advance. Using the sum-table method, these quantities can be efficiently computed using the following formulations:

$$s(i) = F(i) + s(i-1), i \in [i_1, i_M] \tag{5-4}$$

$$s^2(i) = F^2(i) + s^2(i-1), i \in [i_1, i_M] \tag{5-5}$$

with *s(i)* = *s²(i)* = *0* when $i < i_1$. The energy of the reference image can be directly calculated as

$$\sum_{i=1}^{I+N} F^2(i) = s^2(I+N) - s^2(I-1). \tag{5-6}$$

Then the number of operations for computing the $\sqrt{\sum_{i=1}^{N}[f(i-I)-\bar{f}]^2}$ can be reduced to some additions under this feature. The construction of the sum-table takes approximately *3M* operations. This number is much less than the computational cost of the NCC, which is about *3N(M - N)*.

B. Numerator

After solving the problem of denominator, the next process is to find a way to reduce the computational time of the numerator. First we use $T(i) \equiv t(i) - \bar{t}$ and $F(i-I) \equiv f(i-I) - \bar{f}$ to rebuild the formula of the numerator:

$$numerator(I) = \sum_{i=1}^{N} T(i)F(i-I). \tag{5-7}$$

This equation is similar to the definition of the convolution of two images. Therefore, (5-7) can be transferred to (5-8):

$$numerator(I) = T(i) * F(i). \tag{5-8}$$

Based on the convolution theorem, the Fourier transform of (5-8) can be converted to the multiplication of the Fourier transform of these two images, which is shown in (5-9).

$$\mathcal{F}(numerator) = \mathcal{F}(T) \times \mathcal{F}(F). \tag{5-9}$$

Then we can use the inverse Fourier transform to get the value of the *numerator(I)* at each sample index.

The benefit of treating the numerator as a convolution is that we can apply the FFT algorithm to reduce the complexity of the direct 'spatial' computation in (5-7) to *9MlogM* addition/subtraction and *6MlogM* multiplication. This method works faster when *N* approaches *M* and both *M* and *N* are large enough. In [7], Chandrasekaran proved that when *N* = 200, the average performance of the correlation algorithm can be optimized.

Therefore, in this application, we set $N = 200$ and $M - N$ varies in [6, 90] rely on the speed scale chosen for comparison. The superiority of FNCC needs to be verified.

5.5   Experiments and Evaluations

In this section, we study the ability of FNCC in reducing computational cost for providing matching service and compare its performance to the PMCC matching algorithms.

5.5.1   Set Up and Methodology

In this research, we used the same field test data collected in Chandrasekaran's research [7]. In order to derive speed information from the cellular probe's physical parameters, the received signal strength (RSS) of monitored handsets was selected due to its characteristics, such as common availability, economical efficiency, and compatibility. In Chandrasekaran's project, they used GPS enabled HTC Typhoon phones running the Intel-POLS [110] software on the AT&T network and the T-Mobile network to record RSS. A Holux GPSlim236 GPS receiver was used to log the ground truth location information which can be applied to pair the signal tracing from the handset. The signal is recorded every second.

In [7], it focused on evaluating the accuracy of the estimated speed using the proposed correlation algorithm in different situations, such as highway, local arterial, and constant speed trace. Based on previous mathematical analysis of the PMCC and the FNCC in section 5.2 and 5.3, we learned that the FNCC can achieve the same accurate performance in template matching as the PMCC method. This was also proved by our experiments which compared 100 samples for 16 speed traces, as shown in Table 5.1.

Thus, in this research, we concentrated our efforts on evaluating the computational cost of these two approaches in the matching phase. We selected the constant speed trace, which was taken from a driving on a 5 mile stretch of road to simplify the experiment.

Table 5.1: Speed and Searching Range Matching List

5mph	90s	10mph	46s	15mph	30s	20mph	24s
25mph	18s	30mph	16s	35mph	14s	40mph	12s
45mph	10s	50mph	10s	55mph	10s	60mph	8s
65mph	8s	70mph	8s	75mph	6s	80mph	6s

The experiment was processed on our ThinkPad T61 laptop. The detail parameters of the testbed are: Intel Core(TM)2 Duo CPU 2.10GHz, 2GB RAM, 80GB Hard drive. Matlab was used to implement both the PMCC and the FNCC. For each speed scale, we randomly chose 100 samples with 200s length as the testing trace. The searching range is selected from Table 5.1. The location difference is randomly chosen from [0m, 100m] for the urban scenario based on the conclusion that the median localization error is around 96m [7].

5.5.2   Searching Range Influence

Due to the requirement of M and N, we use the field test data to evaluate the performance of the FNCC method and the PMCC method. We selected the speed scale from 5mph to 80mph steps by 5mph.

As shown in Figs.5.4 (a),(b),(c), the computational cost of the PMCC is much larger than the FNCC for all speed scale from 5mph, which has the largest searching range ($S = 90s$) to 80mph, which has the least searching range ($S = 6s$). This difference is shrunken when the speed increased. When we compare the computational cost at 80mph, which

has only 6*s* searching range, we observed that the PMCC is most close to the FNCC but still consumed more than 6 times of the calculation period.
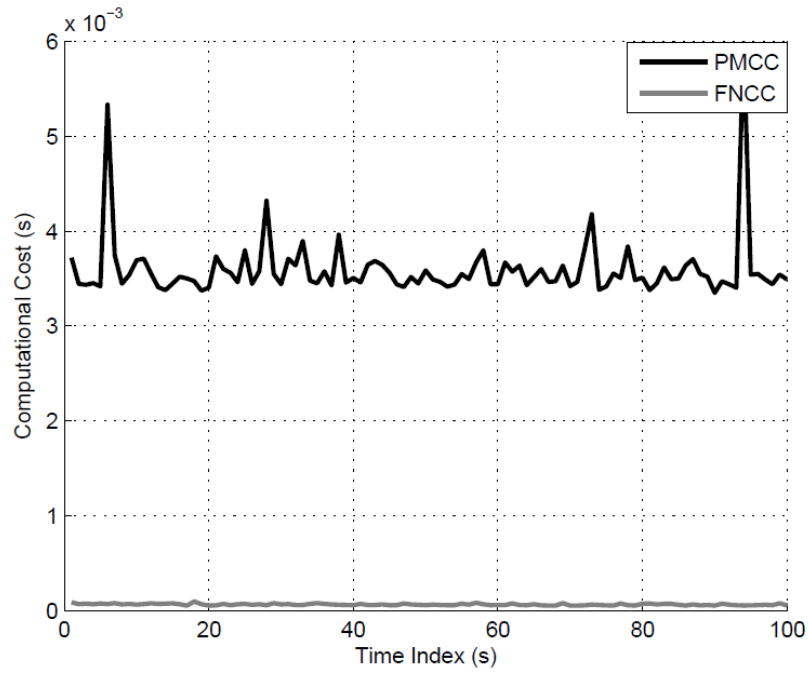
(a) 5mph
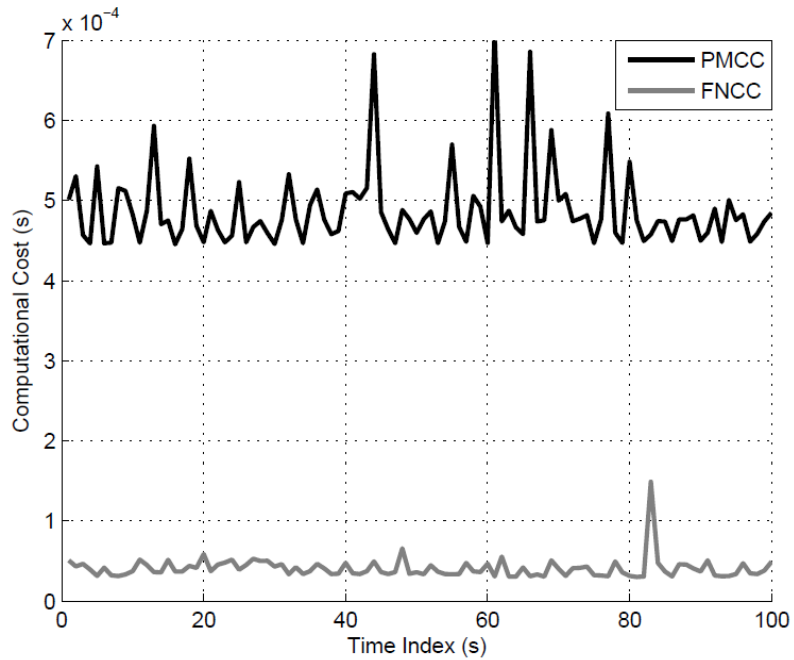
(b) 40mph
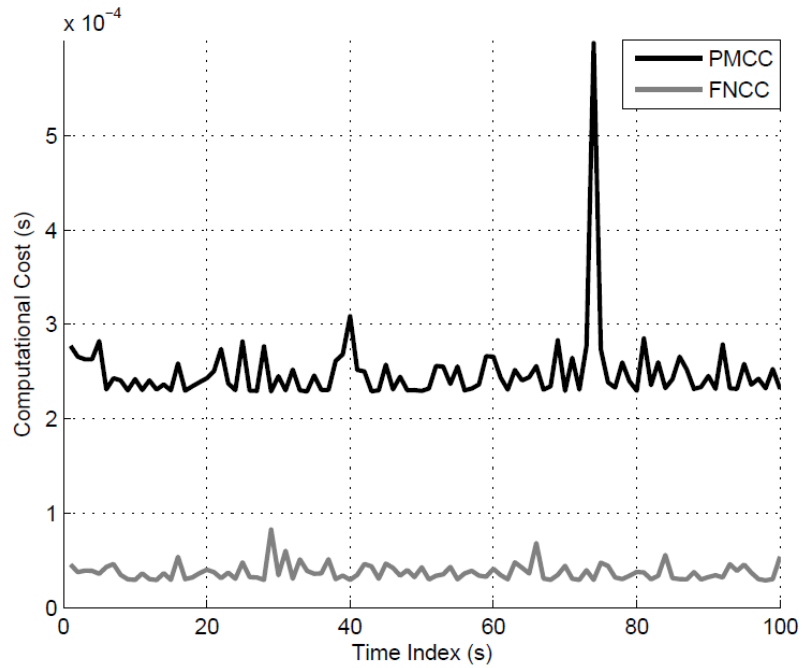
(c) 80mph

Figure 5.4: Comparison of computational time of PMCC and FNCC at different

searching range: (a), (b), (c).
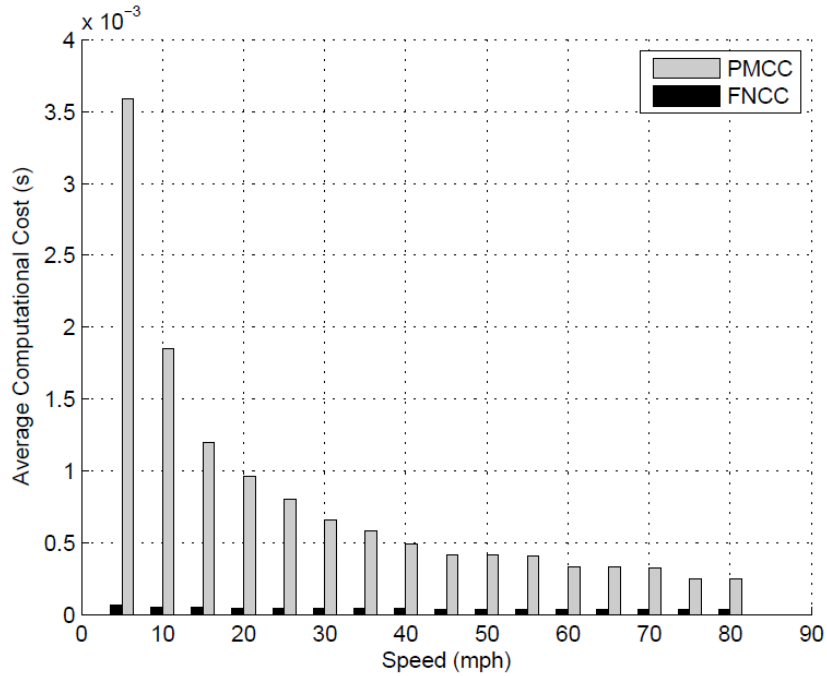
Figure 5.5: Comparison of average computational cost of PMCC and FNCC vary

with speed.

Then we compared the average computation time of 100 tests for each speed, which is illustrated in Fig.5.5. It can be seen that the FNCC provides stable performance due to its operation time, which is dominated by the numerator that has a time complexity of *O(MlogM)*. Contrarily, the computation time of the PMCC is hugely influenced by the value of the searching range **S**. This result matches the mathematical analysis result in Section 5.3.

5.5.3   Ability in Reducing Computational Cost

Table 5.2: Total Time Consumption for Two Algorithms

	PMCC (s)	FNCC (s)	Efficiency
Total Time Consumption	0.1206	0.0034	34.0640

Table 5.2 shows the average operating time of the matching phase for 100 tests, which includes the comparison of the test trace with 80 scaled training traces. The traditional PMCC method consumed about 0.1206s to finish the matching process. The FNCC method, which uses the sum-table and the FFT techniques, can distinctively reduce the computational cost to only 0.0034s for finding the highest correlated training trace. The improvement of the total efficiency is over 34 times, which means that about 97% calculation time can be saved. This is very important, especially when we expect to use centralized method to estimate the real-time speed of each route using cell phone probes.

5.6   Summary

In this chapter, we concentrate our efforts on improving the efficiency of the matching phase of the correlation algorithm, which is an important extension of

Chandrasekaran's work [7]. Based on our survey, this is the first research focused on solving the computational efficiency problem of the correlation algorithm based cellular probe's speed estimation. In order to reduce the time complexity of the matching phase, we first convert this problem into a 1-D curve matching problem. The testing trace is used as the template curve. The training trace that starts from the estimated position at the same road combined with the searching range as the reference curve. A fast normalized cross-correlation algorithm is introduced and evaluated. We present details of the FNCC and how it constructs the sum-table to save the computing time. Our results show that the FNCC can improve the efficiency of the matching phase by about 34 times in urban areas. In other words, using the FNCC method, more than 97% computational period can be saved by reducing the redundant operations in both the numerator and the denominator. Additionally, from Fig.5.5, we can learn that the length of searching range has a positive influence in the computational cost. This illustrates that the FNCC can have steady performance in urban, suburban, and rural environments.

CHAPTER 6

CONCLUSIONS AND FUTURE WORKS

6.1 Conclusions

This dissertation presents a novel congestion mitigating system, challenging the conventional dynamic route guidance system (DRGS) [74] that weighted the directly predicted travel time to get optimal path. The major contributions made by this dissertation are summarized as the following:

1. This study extends the empirical prediction model by investigating the influence of real-time traffic flow fluctuations that were caused by congestions and incidents that happened in related arcs to improve the quality of the forecasted travel time. The causation of congestion in an arc is analyzed by implementing the multi-lane congestion analysis mechanism. The list of related arcs and their real-time traffic conditions are calculated and sent to the iterative travel time estimation model to do a loop prediction process relying on the SVR predictor [79]. The aggregation of SVR, the multi-lane congestion analysis mechanism, and the iterative travel time estimation model represses the fluctuation of adjacent traffic flow inflict by unpredictable incidents and traffic behaviors.

2. This dissertation compares the performance of the traditional coordinates based empirical learning algorithm [2] with the distance based learning algorithm in localizing mobile terminals in outdoor cellular networks. The distance based approach shows superior performance in reducing the errors that exist in the location related parameters and in improving the efficiency of the utilization of geometrical features. The upper

bound of the expect risk of distance errors are mitigated relying on the historical training data set. After calculating the joints of these distances through the trilateration algorithm [10], a direction filter is developed to clean the remaining data set. The upper bound of data clean enhanced method illustrates that using distance instead of coordinates can provide significant improvement in the empirical location estimation algorithm.

3. Finally, this dissertation extends the work of Chandrasekaran [7] by utilizing the 1-D curve matching technique [99] and the fast normalized cross-correlation algorithm [8]. The major contribution achieved in this research is the minimization of 97% of the computational time of the matching phase without debasing the accuracy level of the estimated correlation. This prominent achievement paves a path to provide high quality real-time traffic data for large scale traffic infrastructure.

6.2 Future Works

The primary highlight of this dissertation includes three research areas: predicted travel time based congestion avoidance, data clean scheme enhanced localization, and efficient correlation estimation.

The real-time congestion mitigating system designed in this study exploits our AMPRFP algorithm to improve the traffic efficiency of the urban traffic network. However, the predicted travel time based on AMPRFP algorithm has a potential for experiencing a delayed response to the change of the actual travel time in congestion involved scenarios. This is due to the time required for the analysis of reasons for congestion after it occurs. Additionally, the highway patrol and emergency responders need extra time to verify the report of incidents using traffic cameras. This may be solved

by employing new traffic monitor systems in the urban network that can provide nearly real-time incident detection services.

In order to support this real-time congestion mitigating system, some assistant models which are important in further improving the efficiency and the applicability of this system need to be developed in the future.

The traffic light control plays an important role in deciding the estimated or predicted travel time. In this research, in order to simplify calculation, we assume all traffic lights in the simulated area use fixed time control strategy. Nevertheless, along with the development of ITS, the dynamic control strategy that can cooperate with real-time traffic conditions to adjust the timing and phasing of traffic signals is widely employed in urban transport infrastructure. This advanced control strategy is able to minimize the delay of vehicles and pedestrians going through the intersection. Therefore, one future work of this project is to consider the incorporation of our real-time route guidance system with the dynamic control strategy.

AMPRFP algorithm relies heavily on the detailed moving status of an individual vehicle, especially the leading vehicle. This data is critical to identify the cause of a congestion happened in a route. However, the collection of this kind of information is difficult due to the lacking of sharing mechanisms between vehicles. One possible solution for this problem is the implementation of the Vehicular Ad-Hoc Network (VANET) in the urban traffic network. Vehicles involved in VANET will periodically broadcast their driving status, such as driving direction, driving speed, and current location, through the dedicated short-range communications (DSRC) channel. This

information can be used or relayed by other automobiles and road side units (RSUs). In order to make VANET practical, many technical issues need to be resolved.

Additionally, the development of new techniques for collecting real-time traffic data for local arterial is necessary due to their contributions to the quality of forecasted travel time. In this dissertation, we demonstrate the importance of the cellular network in supporting real-time traffic data collection service. The accuracy of the estimated position of the cellular probe is critical to the performance of cellular probe based FCD. In chapter 4, we illustrated the upper bound that can be achieved using the data clean enhanced learning algorithm. It is obvious that our DCSEEL algorithm can be improved by deeply exploring the position information concealed in each joint. Furthermore, the influence of missing data situation to the DCSEEL algorithm, which is common in the outdoor cellular environment, needs to be investigated.

Although the FNCC can significantly reduce the computational cost in the matching phase, the mechanism that requires comparing the input signal profile with 80 speed scale training traces is quite inefficient in practice. Therefore, to design a 1-D curve matching algorithm that can ignore the scale parameter in the temporal domain is essential in raising the performance of the matching process.

REFERENCES

1.   Sohn, K. and K. Hwang, *Space-based passing time estimation on a freeway using cell phones as traffic probes.* IEEE Transactions on Intelligent Transportation Systems, 2008. **9**.

2.   Wu, Z.L., et al., *Location estimation via support vector regression.* Ieee Transactions on Mobile Computing, 2007. **6**(3): p. 311-321.

3.   Al-Jazzar, S., M. Ghogho, and D. McLernon, *A Joint TOA/AOA Constrained Minimization Method for Locating Wireless Devices in Non-Line-of-Sight Environment.* IEEE Transactions on Vehicular Technology, 2009. **58**: p. 5.

4.   Türkyılmaz, O., et al., *Environment-Aware Location Estimation in Cellular Networks.* EURASIP Journal on Advances in Signal Processing, 2008. **2008**.

5.   Khalaf-Allah, M., *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, 2008. **2008**.

6.   Zhu, H., et al., *Local Positioning with Artificial Neural Network and Time of Arrival Technique*, in *International Conference on Innovative Computing Information and Control 3rd*. 2008.

7.   Chandrasekaran, G., et al. *Vehicular speed estimation using received signal strength from mobile phones*. in *the 12th ACM international conference on Ubiquitous computing   Ubicomp '10*. 2010. New York, NY, USA.

8.   Lewis, J.P., *Fast Normalized Cross-Correlation.* Vision Interface, 1995: p. 120-123.

9.   Phillips, T., C. GauthierDickey, and R. Thurimella, *Using Transitivity to Increase the Accuracy of Sample-Based Pearson Correlation Coefficients.* Lecture Notes in Computer Science, 2010. **6263**(2010): p. 157-171.

10.   Gezici, S., *A Survey on Wireless Position Estimation.* Journal of Wireless Personal Communications, 2007. **44**: p. 20.

11.   Walke, B., *Mobile radio networks : networking, protocols, and traffic performance*. 2nd ed. 2002, New York: Wiley. xxiv, 1152 p.

12.   Pahlavan, K. and A.H. Levesque, *Wireless information networks*. 2nd ed. Wiley series in telecommunications and signal processing. 2005, Hoboken, NJ: John Wiley. xv, 722 p.

13. Wann, C.D. and M.H. Lin. *Location estimation with data fusion for wireless location systems*. in *IEEE International Conference on Networking, Sensing and Control*. 2004. Taipei, Taiwan.

14. Okumura, Y., et al., *Field strength and its variability in VHF and UHF land mobile service*. Rev. Elec. Comm. Lab., 1968. **16**: p. 49.

15. Willassen, S.Y., *A Method of implementing Mobile Station Location in GSM*, in *Information Security* 1998, Norwegian University of Technology and Science.

16. Kan, K.K.H., S.K.C. Chan, and J.K. Ng. *A Dual-Channel Location Estimation System for providing Location Services based on the GPS and GSM Networks*. in *The International Conference on Advanced Information Networking and Applications (AINA)*. 2003.

17. Venkatraman, S. and J. Caffery Jr. *Hybrid TOA/AOA techniques for mobile location in non-line-of-sight environments*. in *IEEE Wireless Communications and Networking Conference (WCNC' 04)*. 2004. Atlanta, Ga, USA.

18. Caffery, J.J., *Wireless location in CDMA cellular radio systems*. The Kluwer international series in engineering and computer science. 2000, Boston: Kluwer Academic. xiii, 189 p.

19. Lindsey, W.C. and M.K. Simon, *Telecommunication systems engineering*. 1991, New York: Dover Publications. xviii, 574 p.

20. Turin, G., W. Jewell, and T. Johnston, *Simulation of Urban Vehicle-Monitoring Systems*. IEEE Trans. Vehic. Tech., 1972. **VT-21**: p. 8.

21. Hashemi, H. *Pulse Ranging Radiolocation Technique and Its Application to Channel Assignment in Digital Cellular Radio*. in *IEEE Vehic. Tech. Conf*. 1991.

22. Bakhtiar, Q.A., N. Pissinou, and K. Makki, *Belief based Data Cleaning for Wireless Sensor Networks*. Journal of Wireless Communications, 2010.

23. Goud, P., A. Sesay, and M. Fattouche. *A Spread Spectrum Radiolocation Technique and Its Application to Cellular Radio*. in *IEEE Pacific Rim Conf. Commun., Comp. and Signal Processing*. 1991.

24. Caffery, J.J. and G. Stüber. *Vehicle Location and Tracking for IVHS in CDMA Microcells*. in *IEEE PIMRC*. 1994.

25. Caffery, J.J. and G.L. Stüber, *Subscriber location in CDMA cellular networks*. IEEE Transactions on Vehicular Technology, 1998. **47**(2): p. 11.

26.   Aatique, M., *Evaluation of TDOA techniques for position location in CDMA*. 1997, Virginia Polytechnic Institute and State University.

27.   Knapp, C. and G. Carter, *The generalized correlation method for estimation of time delay*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1976. **24**: p. 8.

28.   Aatique, M., G.A. Mizusawa, and B.D. Woerner. *Performance of Hyperbolic Position Location Techniques in Code Division Multiple Access*. in *9th Annual Conf. On Wireless Comm. T.R Labs and IEEE Canada*. 1997.

29.   Cong, L. and W. Zhuang, *Hybrid TDOA/AOA mobile user location for wide-band CDMA cellular systems*. IEEE Transactions on Wireless Communications, 2002. **1**: p. 9.

30.   Reza, R.I., *Data Fusion for Improved TOA/TDOA Position Determination in Wireless Systems*. 2000, Virginia Tech.

31.   Catovic, A. and Z. Sahinoglu, *The Cramer-Rao bounds of hybrid TOA/RSS and TDOA/RSS location estimation schemes*. IEEE Communications Letters, 2004. **8**: p. 3.

32.   Nerguizian, C., C. Despins, and S. Affes. *Framework for indoor geolocation using an intelligent system*. in *3rd IEEE workshop on wireless LANs*. 2001. Newton, MA.

33.   Althaus, F. and F.W. Troesch, A. *Uwb Geo-Regioning in Rich Multipath Environment*. in *IEEE vehicular technology conference (VTC 2005 Fall)*. 2005. Dallas, TX.

34.   Ganapathy, V., et al., *An Epistemic Event Based Correlation Approach for Managing Pervasive Networks*. International Journal of Network Management, 2011.

35.   Silventoinen, M.I. and T. Rantalainen, *Mobile Station Emergency Locating in GSM*, in *IEEE International Conference on Personal Wireless Communications*. 1996: India.

36.   Caffery, J.L. and G.L. Stüber. *Radio Location in Urban CDMA Microcells*. in *Personal, Indoor and Mobile Radio Communications (PIMRC' 95)*. 1995.

37.   Silventoinen, M.I. and T. Rantalainen, *Mobile Station Locating in GSM*, in *IEEE Wireless Communication System Symposium*. 1995: Long Island, NY.

38. Wylie, M.P. and J. Holtzman, *The non-line of sight problem in mobile location estimation*, in *5th IEEE International Conference on Universal Personal Communications*. 1996.

39. Xiong, L. *A selective model to suppress NLOS signals in angle-of-arrival (AOA) location estimation*. in *9th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*. 1998.

40. Bakhtiar, Q.A. and N.M. Pissinou, K, *Identification and Validation of Spatio-Temporal Associations in Wireless Sensor Networks*, in *Sensorcom*. 2009.

41. Bakhtiar, Q.A., N. Pissinou, and K. Makki, *Approximate Replication of Data Using Adaptive Filters in Wireless Sensor Networks*, in *ISWPC 2008*. 2008: Greece.

42. Bakhtiar, Q.A., N. Pissinou, and K. Makki, *Estimated Replication of Data in Wireless Sensor Networks*, in *COMSWARE 2008*. 2008: India.

43. Chen, P.C., *A non-line-of-sight error mitigation algorithm in location estimation*, in *IEEE Wireless Communications and Networking Conference*. 1999.

44. McGuire, M., K.N. Plataniotis, and A.N. Venetsanopoulos, *Robust estimation of mobile terminal position*. Electron. Lett., 2000. **36**(16): p. 3.

45. Wylie-Green, M.P. and S.S. Wang. *Robust Range Estimation in the Presence of the Non-line-of-sight Error*. in *IEEE 54th VTS Fall VTC 2001*. 2001.

46. Caffery, J., J. . *A New Approach to the Geometry of TOA Location*. in *IEEE Vehicular Technology Conference (Fall VTC 2000)*. 2000. Boston, Mass.

47. Brookner, E., *Tracking and Kalman filtering made easy*. 1998, New York: Wiley. xxii, 477 p.

48. Bakhtiar, Q.A. and A. Zerguine, *A Generalized Dual Mode Blind Equalization Scheme with Carrier Recovery*, in *EUSIPCO 2005*. 2005: Turkey.

49. Thomas, N.J., D.G.M. Cruickshank, and D.I. Laurenson. *A Robust Location Estimator Architecture with Biased Kalman Filtering of TOA Data for Wireless Systems*. in *IEEE Sypm. on Spread Spectrum Tech. & Appli*. 2000.

50. Venkatraman, S., J. Caffery, and H.R. You. *Location using LOS range estimation in NLOS environments*. in *IEEE Vehicular Technology Conference*. 2002.

51.	Venkatraman, S., J. Caffery, and H.R. You, *A novel ToA location algorithm using LoS range estimation for NLoS environments.* IEEE Transactions on Vehicular Technology, 2004. **53**(5): p. 10.

52.	Al-Jazzar, S., J. Caffery, and H.R. You. *A scattering model based approach to NLOS mitigation in TOA location systems.* in *IEEE Vehicular Technology Conf. (VTC' 02).* 2002.

53.	Al-Jazzar, S. and J. Caffery. *ML and Bayesian TOA location estimators for NLOS environments.* in *IEEE Vehicular Technology Conf. (VTC' 02).* 2002.

54.	He, Y., H.Y. Hu, and S. Zhou, *A TOA Based Believable Factor Mobile Location Algorithm.* Journal of Electronics and Information Technology, 2005. **27**: p. 4.

55.	Shi, M.Y., et al., *A Believable Factor and Kalman Filtering Based Mobile Localization Algorithm*, in *4th IEEE Int'l Conf. on Circuits and Systems for Communications (ICCSC 2008).* 2008: Shanghai. p. 5.

56.	Bahl, P. and V.N. Padmanabhan. *RADAR: An in-building RF-based user location and tracking system.* in *IEEE INFOCOM 2000.* 2000. Tel Aviv , Israel.

57.	Youssef, M.A., et al., *A probabilistic clustering-based indoor location determination system.* 2002, University of Maryland Computer Science Department.

58.	Battiti, R., A. Villani, and T.L. Nhat. *Neural network models for intelligent networks: deriving the location from signal patterns.* in *AINS2002.* 2002.

59.	McGuire, M., K.N. Plataniotis, and A.N. Venetsanopoulos, *Location of mobile terminals using time measurements and survey points.* IEEE Transactions on Vehicular Technology, 2003. **52**(4): p. 999-1011.

60.	Gezici, S., H. Kobayashi, and H.V. Poor. *A new approach to mobile position tracking.* in *IEEE Sarnoff Symposium on Advances in Wired and Wireless Communications.* 2003. Ewing, NJ.

61.	Vapnik, V.N., *The nature of statistical learning theory.* 1995, New York: Springer. xv, 188 p.

62.	Vapnik, V.N., *An overview of statistical learning theory.* IEEE Trans. Neural Networks, 1999. **10**: p. 12.

63.	Sun, G.L. and W. Guo, *Robust Mobile Geo-Location Algorithm Based on LS-SVM.* IEEE Transactions on vehicular technology, 2005. **54**(3): p. 5.

64.   Haykin, S.S., *Neural networks : a comprehensive foundation*. 2nd ed. 1999, Upper Saddle River, N.J.: Prentice Hall. xxi, 842 p.

65.   Hata, M., *Empirical formula for propagation loss in land mobile radio services.* IEEE Transactions on Vehicular Technology, 1980. **29**(3): p. 9.

66.   Faouzi, N.E., H. Leung, and A. Kurian, *Data fusion in intelligent transportation systems: Progress and challenges– A survey.* Information Fusion, 2010. **12**(1): p. 7.

67.   Schrank, D. and T. Lomax, *The 2007 Urban Mobility Report*, ed. T.T.A.M.U. System. 2007.

68.   Chen, K., K. Makki, and N. Pissinou, *A Real-Time Wireless Route Guidance System for Urban Traffic Management and Its Performance Evaluation*, in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*. 2009: Anchorage, AK p. 5.

69.   Kramascz, T.K., *Metropolitan Freeway System 2009 Congestion Report*. 2010, Metro District Office of Operations and Maintenance and Regional Transportation Management Center.

70.   Zhan, F.B. and C.E. Noon, *Shortest Path Algorithms: An Evaluation using Real Road Networks.* Transportation Science, 1998. **32**(1): p. 9.

71.   Tarko, A. and N. Rouphail, *Travel time fusion in ADVANCE*, in *Pacific Rim TransTech Conference: A Ride into the Future*. 1993: Seattle, Washington.

72.   Bhavsar, P., et al., *A Decision Support System for Predicting Traffic Diversion Impacts across Transportation Networks using Support Vector Regression.* Transportation Research Record: Journal of the Transportation Research Board, 2007.

73.   Messer, C.J., C.L. Dudek, and J.D. Friebele, *Method for predicting travel time and other operational measures in real-time during freeway incident conditions.* Transportation Research Record: Journal of the Transportation Research Board, 1973(461): p. 16.

74.   Kühne, R.D., *Data fusion for DRG systems*, in *IFAC Symposium on Transportation Systems*. 1997: Chania, Greece.

75.   Park, T. and S. Lee, *A Bayesian Approach for Estimating Link Travel Time on Urban Arterial Road Network* in *Computational Science and Its Applications (ICCSA 2004)*. 2004. p. 9.

76. Rice, J. and E. van Zwet, *A Simple and Effective Method for Predicting Travel Times on Freeways.* IEEE Transactions on Intelligent Transportation Systems, 2004. **5**(3): p. 8.

77. van Lint, J.W.C., *Online Learning Solutions for Freeway Travel Time Prediction.* IEEE Transactions on Intelligent Transportation Systems, 2008. **9**(1): p. 10.

78. Park, D.J. and L.R. Rilett, *Forecasting Multiple-period freeway link travel times using modular neural networks.* Transportation Research Record: Journal of the Transportation Research Board, 1998. **1617**(1998): p. 8.

79. Wu, C.H., J.M. Ho, and D.T. Lee, *Travel-Time Prediction With Support Vector Regression.* IEEE Transactions on Intelligent Transportation Systems, 2004. **5**(4): p. 6.

80. Du, J. and M.J. Barth, *Next-Generation Automated Vehicle Location Systems: Positioning at the Lane Level.* IEEE Transactions on Intelligent Transportation Systems, 2008. **9**(1): p. 10.

81. Rose, G., *Mobile Phones as Traffic Probes: Practices, Prospects and Issues.* Transport Reviews, 2006. **26**(3): p. 17.

82. Moustafa, H. and Y. Zhang, *Vehicular networks : techniques, standards, and applications*. 2009, Boca Raton: CRC Press. xvii, 445 p.

83. Schönhof, M. and D. Helbing, *Criticism of three-phase traffic theory* Transportation Research Part B: Methodological, 2009. **43**(7): p. 14.

84. (CDOT), C.D.o.T. *How is Speed Calculated?* 2010; Available from: http://www.cotrip.org/helpSpeedCalculation.htm.

85. Dijkstra, E.W., *A note on two problems in connection with graphs* Numeriche Mathematik, 1959. **1**: p. 3.

86. Chrobok, R., et al. *Three categories of traffic data: Historical, current, and predictive*. in *the 9th Int. Fed. Automatic Control Symp. Control in Transpotation Systems*. 2000.

87. Rüping, S., *mySVM–Manual*. 2000, Computer Science Department, AI Unit, University of Dortmund: Dortmund.

88. Program, C.V.O.a.T.I.M., *Rapid Incident Scene Clearance Annual Report - Fiscal Year 2009/2010*, F.D.o. Transportation, Editor. 2010: Tallahassee. p. 15.

89. Cooper, C.D. and D.K. Keely, *A Florida-Specific, MOBILE6-Based CO Screening Model for Air Quality Analysis of Transportation Projects*, in *CO Florida 2004*. 2004, Civil & Environmental Engineering Department, University of Central Florida. p. 15.

90. INRIX, *2009 Annual Report 100 Most Congested Metro Summary*. 2009, INRIX.

91. INRIX, *Inrix National Traffic Scorecard 2009 Annual Report*. 2009, INRIX.

92. Leduc, G., *Road Traffic Data: Collection Methods and Applications.* Working Papers on Energy, Transport and Climate Change, 2008. **1**: p. 55.

93. Kaplan, E.D. and C. Hegarty, *Understanding GPS : principles and applications*. 2nd ed. Artech House mobile communications series. 2006, Boston: Artech House. xvii, 703 p.

94. LaMarca, A., et al., *Place lab: Device positioning using radio beacons in the wild.* Pervasive Computing, Proceedings, 2005. **3468**: p. 116-133.

95. Constandache, I., R.R. Choudhury, and I. Rhee. *Towards Mobile Phone Localization without War-Driving*. in *Proceedings IEEE INFOCOM 2010* 2010. San Diego, CA.

96. Roos, T., P. Myllymaki, and H. Tirri, *A Statistical Modeling Approach to Location Estimation.* IEEE Transaction of Mobile Computing, 2002. **1**(1): p. 11.

97. Ibrahim, M. and M. Youssef, *CellSense: A Probabilistic RSSI-Based GSM Positioning System*, in *IEEE Global Telecommunications Conference GLOBECOM 2010*. 2010: Miami, FL, USA p. 5.

98. Chen, M.Y., et al., *Practical metropolitan-scale positioning for GSM phones.* Ubicomp 2006: Ubiquitous Computing, Proceedings, 2006. **4206**: p. 225-242.

99. Cui, M., et al., *Curve matching for open 2D curves.* Pattern Recognition Letters, 2009. **30**(1): p. 1-10

100. Gibson, D. and C. Tweedy, *An Advanced Preformed Inductive Loop Sensor*, in *North American Travel Monitoring Exhibition and Conference (NATMEC)*. 1998: Charlotte, North Carolina, USA.

101. Sharma, S.C., *Minimizing Costs of Manual Traffic Counts: Canadian Example*, in *Transportation Research Record 905*. 1983, National Research Council: Washington, DC. p. 7.

102. Skszek, S.L., *"State-of-the-art" Report on Non-Traditional Traffic Counting Methods*. 2001, Arizona Department of Transportation: Phoenix. p. 88.

103. Bahler, S., E. Minge, and J. Kranig, *Field test of nonintrusive traffic detection technologies.* Transportation Research Record: Journal of the Transportation Research Board, 1998. **27**(9-11): p. 349-352.

104. Kühne, R.D., *Freeway control using a dynamic traffic flow model and vehicle reidentification techniques*. 1991, Washington, DC Transportation Research Board. 9.

105. Washburn, S.S. and N.L. Nihan, *Estimating Link Travel Time with the Mobilizer Video Image Tracking System.* Journal of Transportation Engineering, 1999. **125**(1): p. 15-20

106. Hellinga, B. and P. Izadpanah, *An opportunity assessment of wireless monitoring of network-wide road traffic conditions*. 2007, University of Waterloo: Waterloo. p. 1-88.

107. Hellebrandt, M., R. Mathar, and M. Scheibenbogen, *Estimating position and velocity of mobiles in a cellular radio network.* IEEE Transactions on Vehicular Technology, 1997. **46**(1): p. 65 - 71

108. Gundlegard, D. and J.M. Karlsson, *Handover location accuracy for travel time estimation in GSM and UMTS.* Intelligent Transport Systems, IET 2009. **3**(1): p. 87 - 94

109. Sato, S. and H. Iwasaki, *A Skeletal Parallel Framework with Fusion Optimizer for GPGPU Programming* Lecture Notes in Computer Science, 2009. **5904**(2009): p. 79-94.

110. Seattle, I.R. *Privacy observant location system (pols)*. 2005; Available from: http://pols.sourceforge.net/.

VITA

KAI CHEN

February 27, 1982	Born, Shijiazhuang, Hebei, China
2000	B.E. in Telecommunication Engineering Beijing Electronics Science & Technology Institute Beijing, Beijing, China
2008	M.S in Telecommunication Engineering Florida International University Miami, FL, United States

PUBLICATIONS & PRESENTATIONS

K. Chen, K. Makki, and N. Pissinou, "A Real-Time Wireless Route Guidance System for Urban Traffic Management and Its Performance Evaluation," the 70th IEEE Vehicular Technology Conference, (Anchorage, USA), September 2009.

K. Chen and N. Pissinou, "Fast Normalized Cross-Correlation Enhanced Floating Car Data Estimation", the 20th Annual Wireless and Optical Communications Conference, (Newark, USA), April 2011.

K. Chen, N. Pissinou, and K. Makki, "Cellular Network Location Estimation via RSS-based Data Clean Enhanced Scheme", the 16th IEEE Symposium on Computers and Communications, (Kerkyra, Greece), June 2011.

K. Chen, K. Makki, and N. Pissinou, "Congestion Reason Analysis based Real-time Route Guidance System for Multi-lane Roads", International Journal on Artificial Intelligence Tools (IJAIT).