ARAB ACADEMY FOR SCIENCE, TECHNOLOGY AND MARITIME TRANSPORT (AASTMT)

College of Engineering and Technology

Construction and Building Engineering Department

USING TRANSIT VEHICLES AS PROBES FOR TRAVEL TIME ESTIMATION ON URBAN ROADS

By

MOHAMED ASHRAF ELSOKKARY

Teaching Assistant, Construction and Building Engineering Department, College of Engineering,

Arab Academy for Science, Technology and Maritime Transport, Egypt.

A thesis Submitted to AASTMT in Partial
Fulfillment of the Requirements for the award of the degree of

Master of Science

IN

ENVIRONMENTAL ENGINEERING

Supervised by

Dr. Akram Soltan Kotb

Associate Professor of Transportation Engineering, Construction and Building Engineering Department College of Engineering, Arab Academy for Science, Technology and Maritime Transport, Egypt.

Dr. Mohamed El Faramawy El Esawey

Assistant Professor of Transportation Engineering, Public Works Department Faculty of Engineering Ain Shams University, Egypt.

Col. Dr. Ayman Sameer El-Dabaa

Rapporteur of the national council for road safety, Egypt.

2015

DECLARATION

I certify that all the material in this thesis that is not my own work has been identified, and that no material is included for which a degree has previously been conferred on me.

The contents of this thesis reflect my own personal views, and are not necessarily endorsed by the University.

(Signature) …………………………………………….

(Date) 25-Jun-2015

We certify that we have read the present work and that in our opinion it is fully adequate in scope and quality as thesis towards the partial fulfillment of the Master Degree requirements in

Specialization

From

College of Engineering (AASTMT)

Date: 25-Jun-2015

<u>Supervisor (s):</u>

Name: Dr. Akram Soltan Kotb

Position: Associate Professor of Transportation Engineering at the Construction and Building Engineering Department, College of Engineering, Arab Academy for Science, Technology and Maritime Transport.

Signature: ………………………

Name: Dr. Mohamed El Faramawy El Esawey

Position: Assistant Professor of Transportation Engineering, Public Works Department, Faculty of Engineering, Ain Shams University.

Signature: ………………………

Name: Col. Dr. Ayman Sameer El-Dabaa

Position: Rapporteur of the national council for road safety.

Signature: ………………………

<u>Examiners:</u>

Name: Dr. Laila Salah El Din Radwan

Position: Professor of Highways and Airports Engineering, Public Works Department, Faculty of Engineering, Cairo University.

Signature: ………………………

Name: Dr. Khaled Adel El Arabi

Position: Professor of Transportation and Traffic Engineering, Public Works Department, Faculty of Engineering, Ain Shams University.

Signature: ………………………

DEDICATION

I would like to dedicate this research to my wonderful wife, my dedicated parents, my lovely sister, and her husband and all of my sincere friends and colleagues with appreciation for the support, help, prayer, and encouragement that I have received during the years that I have devoted for the master's degree.

ACKNOWLEDGMENTS

First, I am grateful to Allah for aiding and supporting me throughout my carrier and studies, for my deeds are by him and for him to aid my brothers and sisters of humanity.

I would like to thank everybody by their names that supported me through the years and through this research. My family, the big loving and caring home that gives a lot and asks for few. My professors and teachers of ethics and manners, Assoc. Prof. Akram Soltan Kotb, Assistant Prof. Mohamed El Faramawy El Esawey, Col. Dr. Ayman Sameer El-Dabaa, Prof. Abd El Meneim Sanad, Prof. Adel Mahmoud Belal, and many thanks to friends that helped me to excel.

ABSTRACT

Obtaining the near real-time information of travel times is a critical element of most applications of Intelligent Transportation Systems (ITS). Instrumenting the roadway infrastructure with inductance loops, cameras and other sensors to obtain travel time data is very expensive leading to an increased data collection cost. The cost becomes even higher if the covered area is large. Hence, cost-efficient approaches for collecting travel time data are highly desirable. This research proposes two models to estimate travel time on urban roads in Greater Cairo using buses as probes. Travel time data were collected using GPS receivers installed on test vehicles and buses that travel along the same urban routes. The cost of GPS receivers is not very high compared to other instruments and hence can be used for large scale deployments of vehicle probe systems. The data were collected at different periods of the day for a comprehensive evaluation. The travel times of bus and automobile were compared in order to explore similarities and differences between the speed profiles. Statistical regression models were accordingly developed to relate automobile travel times/speeds to bus travel times/speeds. The models showed reasonable estimation accuracy taking into account the chaotic nature of traffic in Greater Cairo in addition to the natural variability of travel times. In the first model, average bus speed was used as the dependent variable while average automobile speed was used as the explanatory variable. In the second model, however, travel times were used instead of speeds. Finally, two linear regression models were developed and were found to be the most successful models in terms of their goodness of fit and predictive ability, which the R^2 is 65.9%, and 69.3% for the two models, respectively. The average estimation error of the model did not exceed 17.6% for each run. The developed models could be thought of as an initial step towards the development of a continuous system for collecting and disseminating traffic information to travelers using transit probes data. If such a system exists, it can be used to help traffic management centers in optimizing the performance of the network, especially during congestions. Further research is still necessary to refine the two models as much as possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Symbols	**Nomenclatures**
ATT	Automobile Travel Times
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural networks
APC	Automatic Passenger Counter
APTS	Advanced Public Transportation Systems
ATIS	Advanced Traveler Information Systems
ATMS	Advanced Traffic Management Systems
AVI	Automatic Vehicle Identification
AVL	Automatic Vehicle Location
BPR	Bureau of Public Roads
BTT	Bus Travel Times
DMI	Distance Measuring Instrument
E	Easting
GIS	Geographic Information System
GPS	Global Positioning System
ITS	Intelligent Transportation systems
K-NN	K- Nearest Neighbor
LPM	License plate matching
MAPE	Mean Absolute Percentage Error
N	Northing

TST	Total Stopping Time
UTM	Universal Transverse Mercator

CHAPTER ONE

INTRODUCTION

1.1 Background

During the 20th century, most of the methods used for solving traffic problems were limited to improving physical infrastructure by constructing bridges, tunnels, extra roads, expanding the existing ones and adding traffic signals. With a rapid increase in urban population and vehicles' ownership, these solutions have become more difficult due to physical space constraints and environmental concerns. Realizing that, government agencies and transportation professionals around the world have shifted their focus to new approaches that rely on traffic information to monitor, manage, and improve the performance of road networks. A set of new information-based systems have been designed to support the new approach which are widely known as Intelligent Transportation Systems (ITS) (Mahmoud, 2009).

ITS can be defined as any system that applies advanced technologies and communications to improve the operational and safety performance of the transportation system. Advanced Traveler Information Systems (ATIS), Advanced Public Transportation Systems (APTS) and Advanced Traffic Management Systems (ATMS) are a few examples of ITS applications.

The application of ATIS has become central for mitigating congestion in many cities. The purpose of an ATIS is to collect, estimate, predict, and disseminate traffic information to drivers in real-time so they can better manage their trips. For example, travel time information will allow road users make informed decisions about route choices by avoiding congestion and incident locations. Yet, travel time information is the most important and most complex traffic parameter to collect and disseminate.

Travel time information is a key performance measure for traffic analysts, public agencies, and planners. It has been used extensively for network-wide analysis and evaluation. There is a growing interest in techniques for collecting reliable travel time data as well as models and methods that can be used to analyze and disseminate travel time information to end-users in real-time.

Instrumenting the roadway infrastructure with inductance loops, cameras and other sensors to obtain travel time information is very expensive leading to an increased data collection cost. The cost becomes even higher if the covered area is large. Hence, cost-efficient approaches for collecting travel time data are highly desirable.

This research aims to use transit vehicles as probes to gather travel time information on some urban roads in Greater Cairo. Of particular interest is to examine the potential for using buses as probe vehicles for getting average route travel times.

1.2 Problem Statement

Using transit vehicle as probes offers a number of advantages, as buses cover a large portion of urban networks, have fixed routes and schedules, and in many cases buses are equipped by tracking devices such as Global Positioning System (GPS) for fleet monitoring and management purposes. These advantages present a good opportunity for using buses as probes as a cost-efficient method to collect real-time travel time data.

A transit vehicle experiences more delays than automobiles due to acceleration/deceleration and stoppage at bus stops (dwelling time), and their relatively lower speeds. Despite the fact that transit vehicles and automobiles have different running behaviors, a relationship can be developed to estimate automobile travel times using transit data.

1.3 Research Objectives and Goal

The objective of this research is to develop a model that can be used to estimate the general (i.e. automobile) travel time on urban roads of Greater Cairo using GPS data obtained from buses. The goal is to apply the estimated travel times to determine the average speeds on the road network and hence give an indication about congestion locations in real-time.

The main objective and goal of this research can be achieved throughout the following secondary objectives:

1. Analyzing the feasibility of using GPS-data for tracking purposes in Greater Cairo,
2. Estimating bus travel times and delays using the raw GPS data,
3. Defining the fundamental factors that cause the travel times of buses and automobiles to be different,
4. Investigating the relationship between buses' travel times and automobile' travel times,
5. Developing travel time estimation models that are based on transit data, and
6. Validating and testing the developed model.

1.4 Thesis Organization

In addition to the current chapter, the thesis includes four more chapters in the following structure:

- Chapter Two provides a theoretical background and a literature review that covers the most important topics relevant to this research. It begins by defining real-life travel time, and its statistical description (average, individual). Then, it distinguishes between travel time measurement, estimation and prediction. Subsequently, it reviews in detail the time measurement methods; direct, indirect. Finally, it looks at the nature of the urban street travel time prediction problem and briefly summarizes the state-of-the-art of current research and practice and then review the previous bus probe studies.

- Chapter Three describes the data collection, reduction and cleaning stages. In general, the travel times of buses and automobiles have to be measured on the same travel route and at the same time. Hence, buses and test vehicles were equipped with handheld GPS data loggers to enable continuous data collection along the survey routes. The collected data included date, time, vehicle position, speed, etc. of the vehicle. Data were obtained during working days at different periods of the day to cover various traffic conditions (e.g. AM, mid-day, off-peak, and PM).

- Chapter Four describes the data analysis and modeling of this study. This is achieved by analyzing the characteristics of bus and automobiles travel times, subsequently, models were developed to estimate average travel time of the automobiles using the travel times of buses. The model was validated by using a subset of the data collected.
- Finally, Chapter Five concludes the research, presents recommendations and describes potential future work.

CHAPTER TWO

BACKGROUND AND LITERATURE REVIEW

2.1 Travel Time Definition

Travel time is defined as the time taken for an object to complete the journey, assuming that there are no external factors. In reality, the effect of the environment is inescapable. There are many definitions for real-life travel time, for example (Schrader, et al., 2004) defined travel time as "the amount of time taken to travel from one point to another along a given path". (Lint, 2004) defined the route travel time as "the time it takes an individual traveler (driver, pedestrian, passenger) to traverse that particular route". It is important to highlight that both the average and individual travel time are the two main measures to describe travel time over a specified route/segment. In real-life applications, average travel time is usually used as a more practical measure because it gives a more realistic description of the ground truth situation (Mahmoud, 2009).

2.2 Travel Time Measurement, Estimation and Prediction.

It is important to understand and differentiate between travel time measurement, estimation and prediction. Measurement includes direct methods that imply measuring travel time from the field under real traffic conditions. Estimation and prediction, on the other hand, can both be described as indirect methods for travel times determination. These indirect methods employ some traffic parameters such as: flow, occupancy and speed which physically affect travel time values (Guin and Laval, 2013). Travel time estimation refers to the calculation of the travel times of realized trips based on known speeds, flows, travel times, or other quantities. Nevertheless, travel time prediction refers to the calculation of the travel times on a particular facility (i.e. link/route) for future traffic conditions. The process operates in an unknown state of traffic conditions (Hao, et al., 2009).

Online and offline techniques are both used in estimation process. Online estimation is also referred to as real-time travel time estimation and is aimed at calculating the travel time according to the current traffic conditions. Therefore, online travel times are dynamic as their values change from one time to another. Offline travel times uses historical traffic data that have already been collected and archived (Feng, et al., 2010). Offline methods are usually used in analytical studies to evaluate transportation system performance and to help in designing and planning future systems. Furthermore, historical travel time are also considered a fundamental element in building travel time prediction systems (Mahmoud, 2009).

Travel time measurement, estimation and prediction are the basis of an ATIS. Therefore, the predictive ability of any ATIS system is highly dependent on how successful these steps are completed.

2.2.1 Travel Time Measurement through Direct Methods

Travel time can be directly measured from the vehicles travelling on the road considering all the effects suffered by the vehicle while moving. Direct methods can be divided into different categories depending on many factors such as: operation and instrumentation. A common classification is road-based and vehicle-based techniques (Martí, 2010).

In road-based techniques, the equipment used for collecting data is installed either on the roadway (e.g. loop detectors), or on the roadside (e.g. radars and video cameras). Whereas, vehicle-based techniques are using manual and computerized instruments inside the vehicle to measure the travel times. Manual instruments can vary from simple tools such as using a clipboard and a stopwatch to more advanced sensors such as Global Positioning System (GPS) receivers or electronic Distance Measuring Instrument (DMI) (Mathew, 2014). An illustration of each of the two techniques is presented below.

1. Road-Based Techniques

In these techniques, the key for travel time data collection is vehicle identification. The methods mainly depend on identifying a vehicle at two control points, recording its arrival time at each one then, calculating the travel time of that vehicle as the difference between the two arrival times (Martí, 2010).

Various techniques have been developed for vehicle identification. They significantly differ in their accuracy and complexity. Among these techniques, License plate matching and signature matching are considered to be the state-of-the-practice.

i. License Plate Matching Techniques

License plate matching (LPM) aims at identifying vehicles by recognizing their license plate characters. LPM works by matching vehicle license plates between two or more consecutive control points. This can be done by many tools: tape recorders, video cameras, portable computers and automatic license plate character recognition (Mathew, 2014).

Several ways can be implemented to benefit from the LPM. The most unsophisticated is the manual one, in which the data are collected by human observers who write down the vehicles' plates characters at two different points or record the characters on an audio tape and then process the data to do the matching and calculate travel time.

Most of the time the two points of interest are distributed over a large area, hence, it is important to overcome this problem by deploying video cameras at the checkpoints to collect the data. To follow this approach, the vehicles plates are videotaped at the points of interest and then the video data are encoded/extracted, either manually or automatically. In the manual method, a human observer matches the license plates at the two points by watching the recoded videos. The method, although offering better accuracy, is time consuming. In automatic methods, on the other hand, a computer program is used for license plate characters recognition. The method could offer many benefits such as time and labour cost saving. Nevertheless, the method is suspect to false recognitions caused by adverse weather conditions and bright sun lights (Chang, et al., 2004).

In summary, LPM techniques are considered a good choice for travel time data collection. Manual LPM methods are manpower and time consuming, which decrease their ability to generate real-time results. Automatic LPM methods, however, are impacted by environmental conditions which make the plates sometimes hard to observe clearly. Furthermore, travel time data are only limited to locations where video cameras are installed (Mathew, 2014).

ii. Signature Based Techniques

Signature-based techniques are used to overcome the limitations of the plate matching techniques by observing other characteristics of the vehicle such as: its color, type and model that are less affected by environmental conditions.

Signatures can utilize a number of point detectors such as: video cameras, inductance loop detectors, laser scanning detectors, electronic toll identifications and vehicles equipped with Bluetooth mobile devices (Barceló, et al., 2010).

One advantage of this technology is its ability to work for long time without being affected by external conditions. As well, it can generate real-time data by communicating directly with traffic centers. On the other hand, the extremely high cost of installation and maintenance is the main disadvantage of this technology (Martí, 2010).

2. Vehicle-Based Techniques

In vehicle-based techniques, the equipment used for data collection is installed in a vehicle which travels between the points of interest. Similar to road-based techniques, several methods can be used for data collection with varying degree of accuracy and complexity. These methods range from test vehicles that gather data manually to Intelligent Transportation System (ITS) probe vehicles equipped with more advanced technology.

In test vehicle techniques an observer sitting inside the vehicle keeps track of the vehicle location and timestamp at each check point, then reports this information to the traffic center either after or during the test runs (Mahmoud, 2009).

Vehicles which are able to collect data with ITS technology and transmit it to a traffic center are called ITS probe vehicles. Such vehicles collect travel time data by using multiple techniques such as: Automatic Vehicle Identification (AVI), or Automatic Vehicle Location (AVL).

i. *Automatic Vehicle Identification (AVI)*

Probe vehicles are equipped with electronic tags. These tags communicate with roadside transceivers to identify unique vehicles and obtain travel times between transceivers as shown in Figure 2.1 (Turner, et al., 1998).

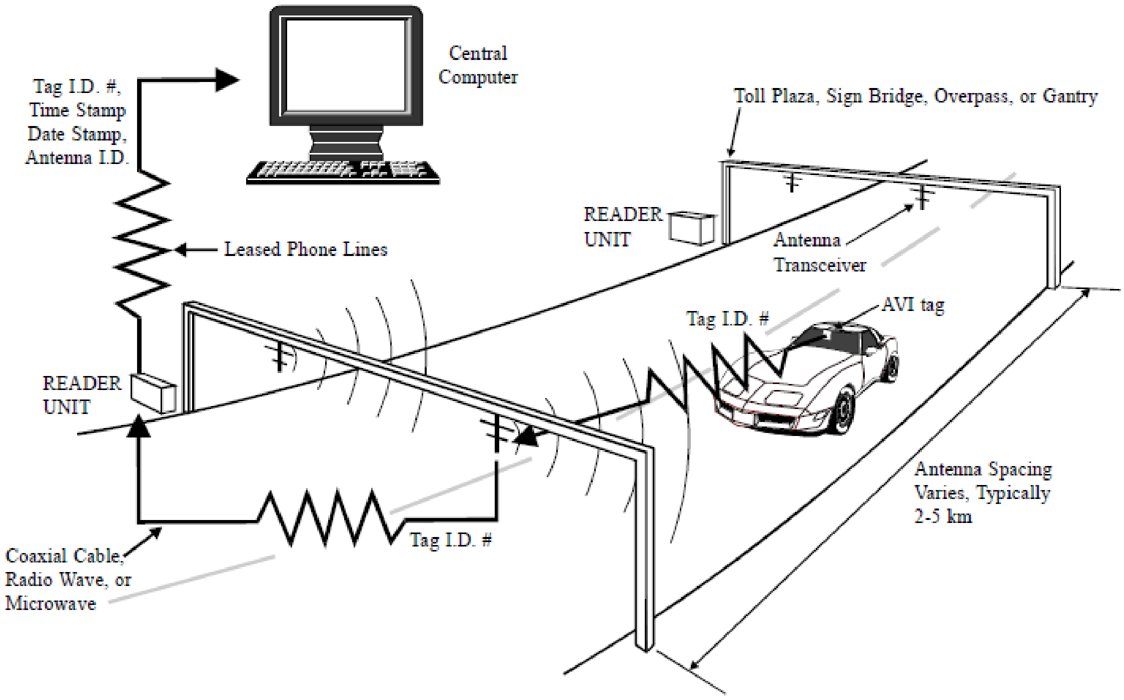

Figure 2.1 AVI Vehicle-to-Roadside Communication Process.

(Turner, et al., 1998)

ii. *Automatic Vehicle Location (AVL)*

This technique has been frequently used by transit agencies. Probe vehicles equipped with GPS receivers and two-way communication can be tracked in real-time and used for estimating useful traffic measures such as travel time as shown in Figure 2.2.

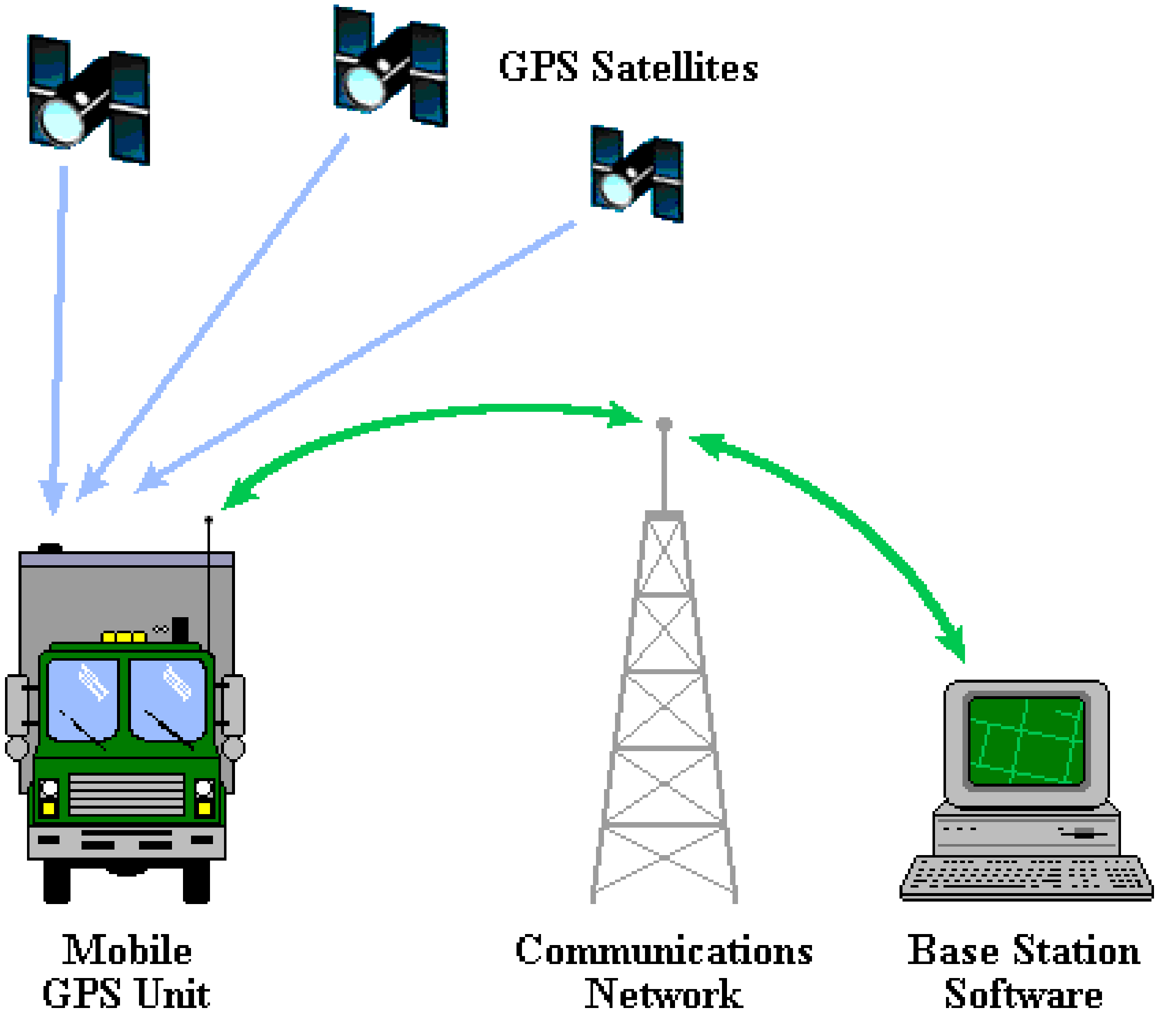

Figure 2.2 AVL Communication Process.

(Sec14)

An advantage of using vehicle-based techniques is their ability to collect data for large regions without interrupting the traffic flow and with minimal cost per unit. On the other hand, the more the number of probe vehicles is, the more accurate the collected travel time is (Jiang, et al., 2006). Accordingly, to get more accurate information, more probe vehicles need to be employed, which means a large increase in both construction cost (purchase necessary equipment, install it, and train personnel to operate the system and collect data) and operating cost.

2.2.2 Indirect Methods

Indirect methods derive travel time as a function of other traffic parameters such as speed, volume, density, etc. They can be classified into three main categories: instantaneous, model-based and data-driven.

1. Instantaneous Methods

These methods are commonly used for online travel time estimation on freeways due to the stability of the traffic conditions (e.g. speed, flow and density) under normal conditions. Despite their success on freeway facilities, these methods are not very successful on arterial roads where the interruption of traffic is natural and thus violates the main stationary traffic conditions assumption.

There are three main techniques used as instantaneous method; the half distance, the average speed method, and the minimum speed method. The half distance approach is the most common method for estimating travel time on highway. It assumes that the entrance speed applies to one half of the link and the exit speed applies to the other half. The second method, the average speed method, assumes that the average of the downstream and the upstream speeds applies to the link. In the last technique, the minimum speed method, the minimum speed between the upstream and downstream stations of a given link is used to calculate the travel times (Mahmoud, 2009).

2. Model-Based Methods

Model based methods estimate travel times using traffic propagation models of traffic flow theory. Different simulation models have been used to estimate travel time. Microscopic models predict travel time directly based on the interactions between vehicles caused by car following, gap acceptance and risk avoidance theories. As for macroscopic models, they derive the travel time indirectly by predicting the characteristics of a traffic stream such as flow, average speed, density and stability properties. The ability of model-based methods to quantify traffic conditions is the main advantage which helps to provide deeper understanding of traffic behavior in the area of interest.

3. Data-Driven Methods

In these methods, travel times are derived from other traffic parameters using different mathematical and statistical relations. Examples include regression methods, time series methods, k- nearest neighbors methods and artificial neural networks which are all based on the usage of historical data to infer present information, depending on the fact that traffic patterns often repeat themselves over time.

i. Regression Methods

Regression analysis is a statistical technique that tries to express output, called the dependent variable, as a function of the input(s), called the independent variable(s) (Alpaydin, 2010). Multiple travel time estimation and prediction methods were developed based on regression analysis. Examples include linear regression with stepwise variables selection technique and more advanced tree-based methods (Kwon, et al., 2000).

(Nikovski, et al., 2005) conducted a comparison between five statistical methods: linear regression, neural networks, regression trees, K-nearest neighbors and locally-weighted regression. The linear regression was shown to be accurate especially for large historical data. It was also shown that linear regression has a high potential for short-term travel time estimation.

ii. Time Series Methods

In statistics, a time series is a sequence of measurements made consecutively. A time series forecasting model is a statistical method that is used to estimate or predict certain measure, by studying its behavior in the past to capture the essential features of the long term behavior of the system (Mahmoud, 2009).

ARIMA, Autoregressive Integrated Moving Average, is a very famous time series model that was the first introduced by (Ahmed and Cook, 1979).

Another well-known time series model is the Kalman Filter which is used to improve travel time estimates, filter the noisy travel time data, and fill in data gaps. For example, (Lint, 2008) proposed a new extended Kalman Filter based on online learning approach to predict travel time. (Liu, et al., 2006) used two distinct ways of using Kalman Filters to address the problem of short-term urban arterial travel time prediction. (Li, 2002) tried to setup a bus real-time prediction model using Kalman Filter theory and the results show that the estimated bus travel time is in agreement with the actual value.

iii. K-Nearest Neighbor Methods

K-Nearest Neighbor (K-NN) is a supervised learning, non-parametric statistical method, in which the value of an object is estimated based on the values of its neighbors in the training space. It assumes that objects close in distance are potentially similar. The method has no predefined assumptions that relate the different input parameters to the output parameter. (Mahmoud, 2009) applied K-NN models to predict travel times for arterial links. The method was shown to achieve the best performance for estimating travel time on highways compared to other methods; average speed, half-distance, minimum speed and average travel time. (El Esawey and Sayed, 2011) used buses as probes for neighbor links travel time estimation. The authors proposed a general framework to integrate historical link travel time data and sparse bus travel time data for travel time estimation on a network. The average error value was found to be 15.4% which was considered acceptable in view of the considerable travel time fluctuations in the study area.

iv. Artificial Neural Networks

An Artificial Neural Network (ANN) is a computational model that is based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes or learns, in a sense based on that input and output. An ANN can perform highly complex mappings on nonlinearly related data by inferring subtle relationships between input and output parameters.

An ANN consists of several layers of interconnecting processing elements named neurons. In addition, activation functions applied with neurons to control the signals passing through the network (Young II, et al., 2008).

An artificial neuron represents a systematic unit correlating input-output or by another word an artificial neuron is a single processing element in an ANN. This single unit is more commonly referred today as a perceptron unit. A single perceptron computes the output of the neuron by forming a linear combination of weights and real-valued input values, where this value is generally translated by a non-linear activation function. For example, the perceptron, shown in Figure 2.3, has a set of independent inputs (X_1, X_2, X_3,…., X_m), connection weights (W_1, W_2, W_3,…., W_m), bias (B), and a dependent variable (Y). For this neuron, the inputs are multiplied with the connection weights and summed together at the node. The bias is also added to this sum and the value is entered into an activation function ($f(\sum(W^T X+B))$. The result of this activation function is equal to the dependent variable. The weights are unknown and represent the strength of the association between independent and dependent variables (Young II, et al., 2008).

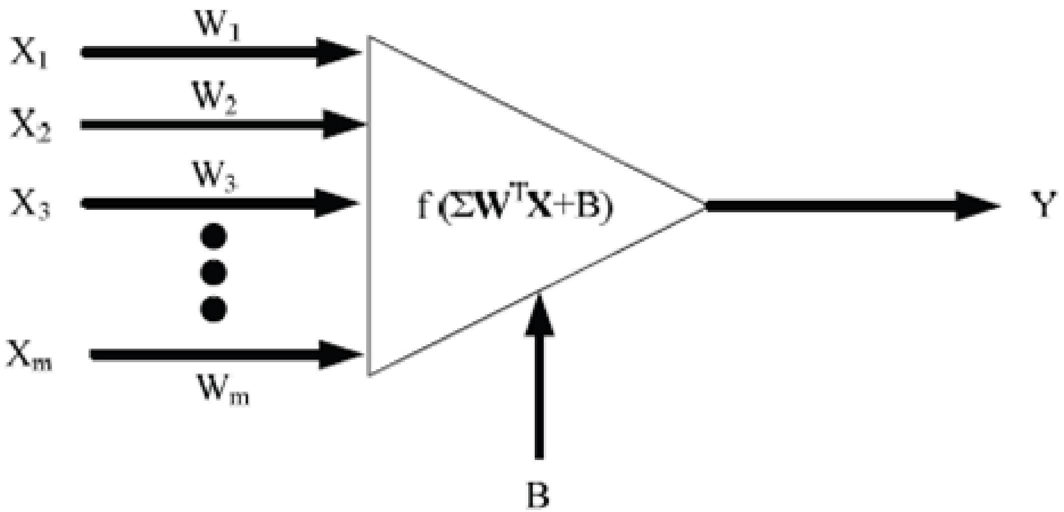

Figure 2.3 Artificial Neuron.

(Young II, et al., 2008)

There are many studies that used ANNs for estimating and predicting travel times. For example, (Mark, et al., 2004) developed an ANN that was capable of forecasting experienced travel time between two points on a freeway section. Computational experiments demonstrated that the developed ANNs were able to reasonably predict experienced travel time. (Farhan, et al., 2002) used empirical data obtained from AVL and APC systems to develop an ANN model to predict bus travel times. (Chien, et al., 2002) computed transit travel times by applying two ANNs, trained by link-based and stop-based data. The reliability analysis showed that enhanced ANNs can perform accurately for both single and multiple stop prediction. Moreover, the stop-based ANN was preferred when there were multiple intersections between stops, while the link-based one was considered more suitable for the pairs of stops with few intersections.

2.3 Urban Travel Time

In the next section, the urban street travel time estimation problem is discussed and the state-of-the-art relevant research and practice is briefly summarized. Subsequently, the previous studies that analyzed the use of buses as probes are analyzed.

2.3.1 Urban Travel Time Estimation and Prediction Problem

Complex traffic environment and limited data sources are the two major problems in urban travel time estimation and prediction. These two problems call for innovative traffic modeling techniques and better traffic detection technologies, which are both in short supply. This conflict between the increasing need and shortage of resources is one of the major causes for the lack of forecasting travel time information for urban streets.

1. Complex Traffic Environment

An urban street is located in an urban area with relatively high density of driveway access, traffic signals and intersections which interrupt traffic flow. In addition, other factors may exist and cause the travel time estimation and prediction for urban streets to be more complicated than freeways. Examples of these factors include, but are not limited to, pedestrians, cyclists, transit buses, on-street parking, and vehicle temporary stop.

2. *Limited Data Sources*

Due to the complex traffic environment, traffic monitoring devices required for urban streets are more than freeways to obtain an acceptable level of accuracy. For example, Sen, et. Al, 1996 concluded that a large number of additional loop detectors are needed for complete monitoring of traffic status on signalized streets (Pu, 2008).

There are two categories of traffic monitoring devices: point detection and route detection. Point detection includes inductive loop detectors, laser sensors, infrared sensors, ultrasonic detectors, remote traffic microwave sensors, video cameras, etc..

The most common traffic detection device in use is inductive loop detectors. Many travel time estimation and prediction studies were based on loop detector data (Sisiopiku, 1994, Anderson and Bell, 1997, Palacharla and Nelson, 1999, Jiang and Zhang, 2001, Stathopoulos and Karlaftis, 2003, Lucas, 2004, Robinson and Polak, 2005 and Guo and Jin, 2006). Despite their popularity, loop detectors have disadvantages when used for ATIS applications especially on signalized urban streets since they are mainly designed for traffic signal control purposes at signalized intersections, where they measure traffic volume and occupancy. Hence, they are not suitable for providing travel time and speed information which are required in an ATIS. (Mahmoud, 2009) analyzed urban travel time estimation using point detectors data and the results showed that the ANN and K- Nearest Neighbor (K-NN) methods significantly outperformed linear regression models.

Route detection deploys the so-called floating cars or probe vehicles. In theory, any vehicle can be a probe as long as the vehicle can be tracked continuously or recognized at the starting and ending points of a route. Examples are personal vehicles instrumented with automatic identification tags (Chien and Kuchipudi, 2003 and Dion and Rakha, 2006), identifiable by video cameras (MacCarley, 1998 and Innamaa, 2001), traceable by cellular phone signals (Bar-Gera, 2007 and Fontaine and Smith, 2007), or equipped with global positioning satellite (GPS) devices (Quiroga and Bullock, 1998 and Taylor, 2000) (Pu, 2008).

2.4 Buses as Probes

2.4.1 Automatic Vehicle Location (AVL) Systems

Early AVL systems in transit buses have been designed for real-time operations and management such as emergency response and computer aid dispatch. In the mid-1990s, archived AVL data and Automatic Passenger Counter (APC) data were used to improve transit performance and management. There are currently two types of AVL systems, archived AVL and real-time AVL.

Archived AVL refers to an AVL system that employs an on-board computer to record and store bus operation data and upload it to a data tank at the end of bus daily operation. Whereas real-time AVL refers to an AVL system that sends bus travel and operation data to a central computer in real-time. Although archived AVL and real-time AVL can also be physically integrated, different data is generated from different functional components (Pu, 2008).

Archived AVL generates location-driven records at specific locations such as a bus stop. It usually contains a good amount of bus operational information. However, only a small amount of this data is useful to infer travel time/speed. Typical collected data include bus travel time, travel distance, stop dwell time, number of stops, and number of boarded/alighted passengers at stops. Dwell time measures the time elapse between the bus door-open time and the door-close time at a stop.

Real-time AVL generates time-driven information including position, speed and heading direction which are the most relevant information for a bus probe study.

Both types of AVL data can provide some travel time and/or speed information that is suitable for bus probes application. However, for real-time applications, the use of real-time AVL data becomes a must. In the past, archived AVL data has been the dominant data source for bus probe studies and practically no real-time AVL data has been used in urban street travel time estimation in ATIS (Pu, 2008).

2.4.2 Past Bus Probe Research

In recent years, the interest in predicting travel times has increased with the increased number of ITS deployments. Some studies utilized loop or other static detector data as mentioned before. Others have studied the use of probe vehicles to estimate link/route travel time. (Du, 2005) developed a geo-statistical model to estimate the average travel time for each individual road link in a network. The author discussed the processing of raw GPS data and converting them into link-by-link data and finally estimating link travel time for the whole road network. The average absolute error for routes travel time estimation was 16.8% which is consistent with the error levels in previous work.

In another study by (Hao, 2013), arterial traffic modeling methods were introduced where the proper transportation domain knowledge (such as traffic flow theories or principles) and advanced machine learning and optimization techniques were explored. Domain knowledge describes the systematic patterns of arterial traffic flow that needs to be respected, while learning and optimization techniques were used to reconstruct such patterns from mobile data and estimate parameters of the patterns when needed. The methods developed by (Hao, 2013) were tested and validated via multiple urban traffic modeling applications and the results showed that: (i) sample queued vehicles contain more useful information than sample free flow vehicles; (ii) one sample queued vehicle is required for a cycle in order to estimate signal performances of the cycle; (iii) the proposed mobile sensor based modeling methods work better for congested traffic conditions than less congested ones.

On the other hand, many researchers investigated the potential use of transit vehicles as probes. (Cathey and Dailey, 2002) described a system that uses transit vehicles as probes to collect travel time and speed data. (Tantiyanugulchai and Bertini, 2003) compared speeds and travel times of probe buses to those obtained from GPS instrumented vehicles. The authors used bus trajectories to obtain the mean travel time and travel speed of buses. Subsequently, they presented two imaginary scenarios of what they call the “hypothetical” bus and the “pseudo” bus. The “hypothetical” bus is a bus that does not stop for passengers (El Esawey and Sayed, 2011). As a result, the travel time of a hypothetical bus is computed as the running time minus the total stopping time at all stops. A “pseudo” bus is a bus that runs with the maximum recorded instantaneous speed between a pair of stops. The results showed that the average test vehicle speed is about 1.66 times greater than real bus speed. While the hypothetical bus speed is equivalent to 1.03 to that of the test vehicle. And the test vehicle speed is about 0.79 of the pseudo bus speed (El Esawey and Sayed, 2011).

(Chakroborty, et al., 2004) compared bus travel times (BTT) with automobile travel time (ATT) and suggested a functional form that predicts ATT based on BTT. Data were collected for both bus travel times and average automobile travel times for five arterial segments during different periods. A simple linear equation was suggested for the conversion. This model had two parts; the first part considered the travel time under free flow and the second part considered the total bus stopping time at all stops (TST). This equation is shown below where (b) is a coefficient that gives the best fit for the given data but fails to give any meaning into the relation between the variations in the data of BTT and ATT.

$$ATTp = \frac{Length\ of\ the\ section}{Free\ flow\ speed} + b(BTT - TST)$$

(Vanajakshi, et al., 2008) discussed the necessary work required for implementing Advanced Public Transportation System (APTS) in Indian traffic conditions using buses as probe vehicles. This study is one of the first approaches to predict travel time under such traffic scenario where GPS data were collected from three consecutive buses traveling in the route corroborating the prediction algorithm based on a Kalman Filter technique.

(Pu, et al., 2009) proposed a generic real-time estimation framework and presented two case studies for examining the sensitivity of bus probes to real-time non-transit vehicles traffic conditions on signalized urban streets. They concluded that the framework represents a possible logical solution of implementing real-time bus probes by utilizing both historical bus-car speed relationships and real-time bus travel information.

(Uno, et al., 2009) used bus probes data to study travel time variability on urban corridors. Acceleration, deceleration, and stopping times are estimated and eliminated from bus travel time to estimate automobile travel times (El Esawey and Sayed, 2011).

(El Esawey and Sayed, 2011) used buses as probes for neighbor links travel time estimation. Neighbor links were defined as nearby links that share similar characteristics and are subject to the same traffic conditions within a road network. The authors proposed a general framework to integrate historical link travel time data and sparse bus travel time data for travel time estimation on a network. The purpose is to estimate travel times on links that are not covered by existing sensors, using their travel time relationships with neighbor links. Neighbor links travel time estimation accuracy, using bus probes data, is assessed using the Mean Absolute Percentage Error (MAPE). The value is 15.4% which is an accepted accuracy level in view of the considerable travel time fluctuations in the study area.

(Gao, et al., 2013) employed a Kalman Filter technique to predict the next bus arrival time using real-time data collected from probe buses. The experimental results showed that the model provided a higher level of veracity and reliability of travel time forecasting in the case of frequently changing traffic conditions.

(Pulugurtha, et al., 2014) examined the relationship between car and bus travel time as most buses operating in urban areas are equipped with AVL units. The role of key influential factors on the ratio between the two travel times was also evaluated and examined to assess the use of buses as probe vehicles. These factors include the number of signalized and non-signalized intersections, the number of driveways, bus-stops, lanes, turns made by the bus and traffic volume. The results show that car travel time can be estimated if bus travel time, the number of signalized intersections per unit distance, the number of lanes, and traffic volume are known.

Using buses as probe vehicles to predict the average automobile travel times on urban corridors in Greater Cairo can be beneficial due to the existence of a large number of buses (about 5,000 buses) which run on the most used arterials (the ones that are of greater importance in terms of average automobile travel time prediction). Generally, they have higher frequencies during peak hour and most buses can be equipped with GPS by the transit agency for predicting bus arrival time. These characteristics of bus routes and schedules make them ideal as probe vehicles.

CHAPTER THREE

DATA COLLECTION AND PREPARATION

3.1 Introduction

Transportation-related data, usually come from different sources: simulation models, network sensors, pilot surveys, etc. While simulation models are usually used to save cost and time, tools and manpower, they always have the drawback of generating indirect real-life data. The data obtained from network sensors have the advantage of being real-life data, but since the researcher has no control over the way the data has been collected, a lot of data cleaning and pre-processing is needed.

Collecting the data locally through a pilot survey has the advantage of generating user-specific data. At the same time, the analyst has control over the test location, time and tools. On the other hand, it is relatively costly in terms of time, manpower and acquiring the tools. Hence, it is up to the researcher to decide which method to use based on the research scope, the budget and the availability of tools and manpower (Mahmoud, 2009). For this research, the data were collected by the researcher.

As mentioned earlier, the main objective of this research is to estimate car travel time on urban roads by using transit vehicles as probes. Therefore GPS data such as date, time, vehicle position, speed and heading angle of the vehicle should be collected for both transit vehicles and test automobiles.

3.2 Selected Roads

In this research, the data were collected on a sample of main urban roads in Greater Cairo for the two opposite directions. Among the roads selected were Gesr El-Suez street, El-Harm street, El-Nasr street, Salah Salem street, Fesal street, and Gameet El-Dewl El-Arabyia street (Figure 3.1to Figure 3.6 show the locations of these roads) which have high traffic demand, high pedestrian density, different traffic behaviors and multiple intersections. Adding to this, it is very important to highlight that transit vehicles cover these roads. Table 3.1 shows the length and the number of intersections of the data collection segments on these urban roads.

Figure 3.1 Gesr El-Suez Street.

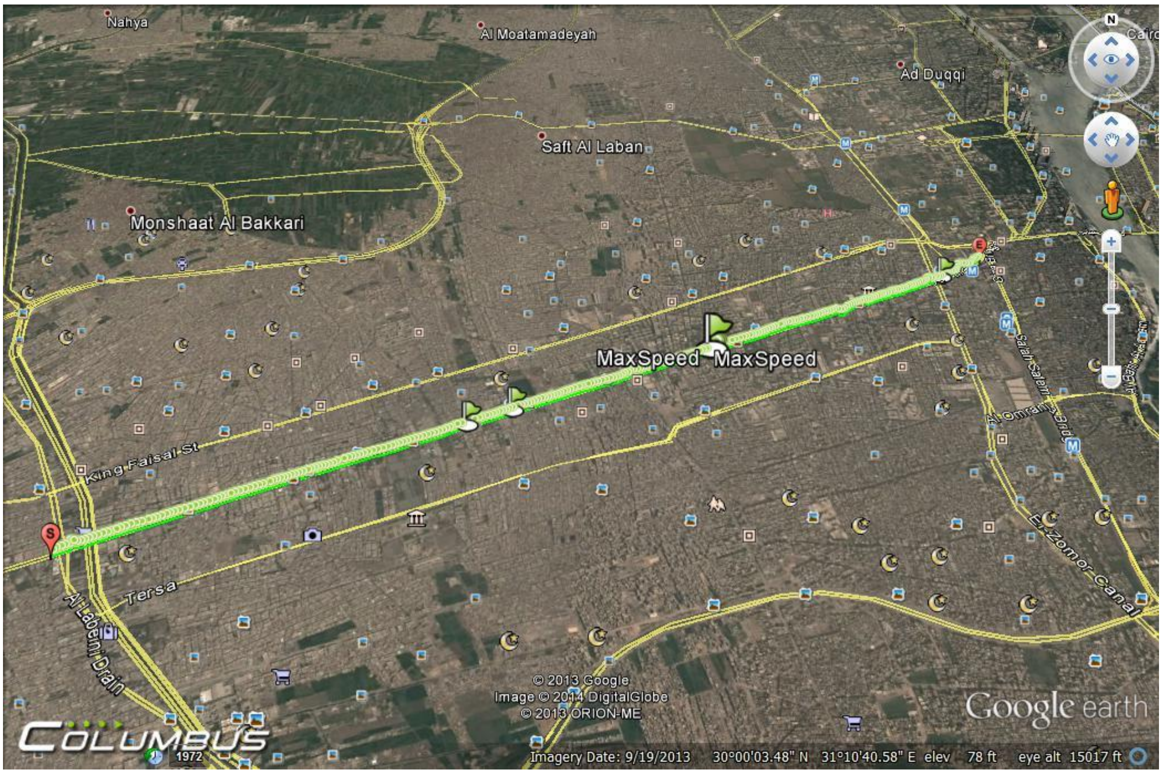

Figure 3.2 El-Harm Street.

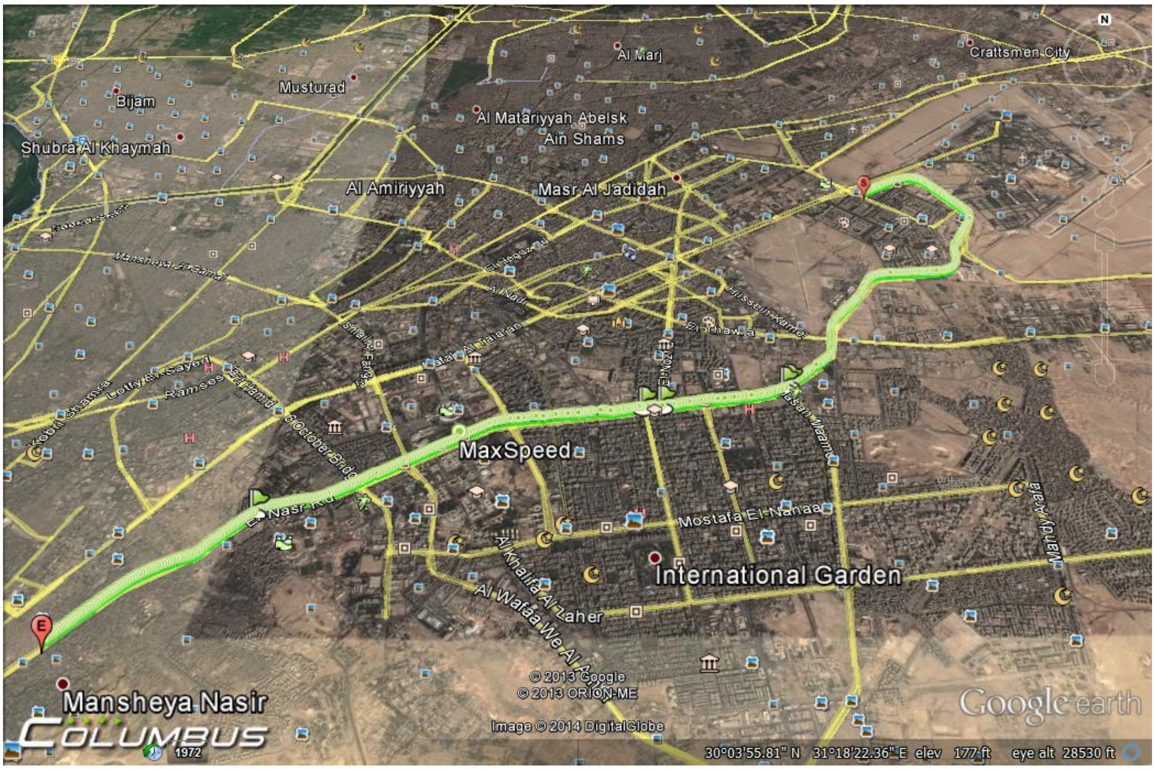

Figure 3.3 El-Nasr Street.

Figure 3.4 Salah Salem Street.

Figure 3.5 Fesal Street.

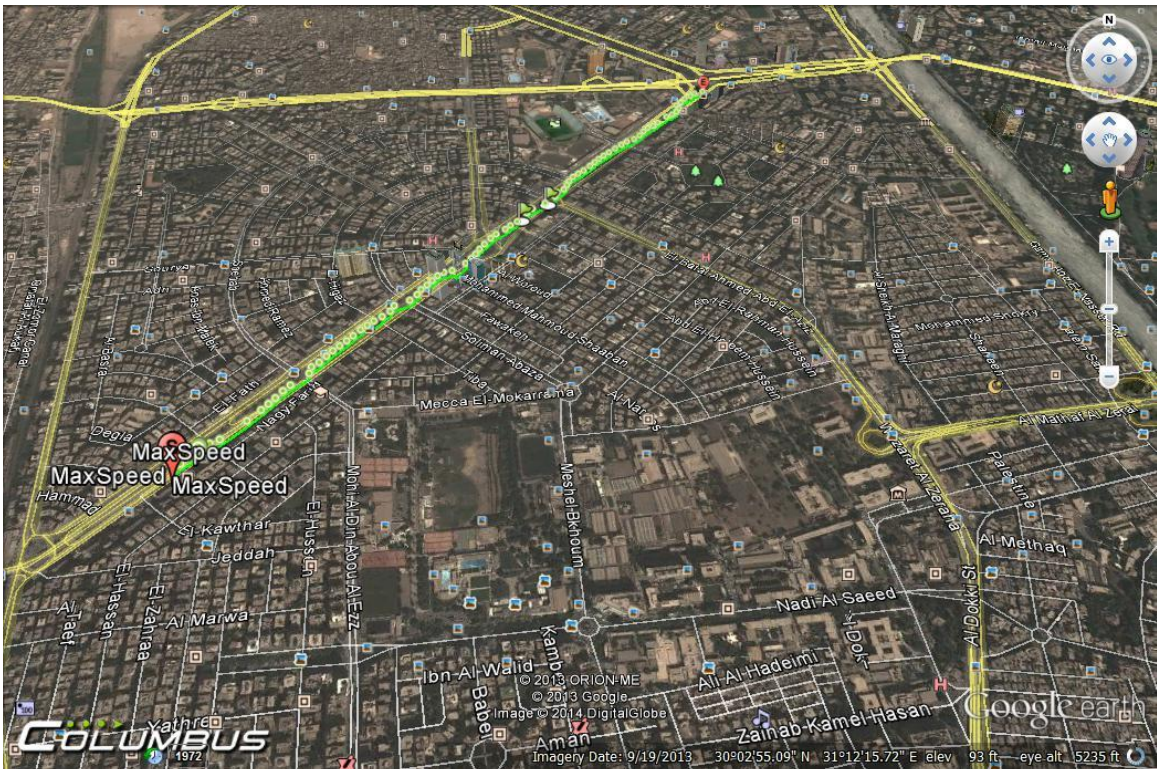

Figure 3.6 Gameet El-Dewl El-Arabyia Street.

Table 3.1 Description of the roads at which data was collected.

Street Name	Length (Km)	No. of intersections
Gesr El-Suez Street (From El-Salam Park to Ebn Sendr Square)	10	5
El-Harm Street (From El-Maryotya Bridge to El-Giza Bridge)	6.5	8
El-Nasr Street (From Sun city to Manshyet El-Bakry)	16	9
Fesal Street (From El-Maryotya Bridge to El-Giza Bridge)	6	6
Salah Salem Street From Military College to El-Abasya Bridge	7	5
Gameet El-Dewl El-Arabyia Street (From El-Mehwar Bridge to El-Sudan Street)	2	7

3.3 Data Collection Period

GPS data were collected at the AM and PM peak periods and also during the off-peak period in normal working days (from Sunday to Thursday) at the end of 2012 and the beginning of 2013. In total, 30 trips were completed for both transit vehicles and cars.

Table 3.2 shows a summary of the thirty data collection trips including the road name and trip direction. It also refers to two groups (G1, G2) who travel on the same street one following the other within at least thirty minutes. Each group consists of three persons; one on the bus with a GPS device to record the travel time, the number of stops at the bus stations and their timing. At the same time, the other two persons were on the car; one as a driver and the other as an observer with a GPS device to record the travel time for the same route noting any incidents along the route. Finally, the table includes the day and the time of the data collection process.

Table 3.2 The day and the time of the data collection process.

No.	Street Name	Direction		Working Day	Time
1	Gesr El-Suez Street	Forward	G1	Monday, November 26, 2012	9:54 AM
2			G2	Monday, November 26, 2012	10:28 AM
3		Backward	G1	Monday, November 26, 2012	11:42 AM
4			G2	Monday, November 26, 2012	12:19 PM
5	El-Harm Street	Forward	G1	Monday, November 26, 2012	3:15 PM
6			G2	Monday, November 26, 2012	3:54 PM
7		Backward	G1	Monday, November 26, 2012	4:00 PM
8			G2	Monday, November 26, 2012	4:30 PM
9	El-Nasr Street	Forward	G1	Tuesday, November 27, 2012	3:38 PM
10			G2	Tuesday, November 27, 2012	4:09 PM
11		Backward	G1	Tuesday, November 27, 2012	5:35 PM
12			G2	Tuesday, November 27, 2012	6:10 PM

13	Fesal Street	Backward	G1	Wednesday, November 28, 2012	6:16 PM
14			G2	Wednesday, November 28, 2012	6:44 PM
15	Salah Salem Street	Forward	G1	Thursday, November 29, 2012	4:18 PM
16			G2	Thursday, November 29, 2012	4:54 PM
17	El-Harm Street	Backward	G1	Sunday, December 02, 2012	8:42 AM
18			G2	Sunday, December 02, 2012	9:19 AM
19	Gameet El-Dewl El-Arabyia Street	Backward	G1	Monday, December 03, 2012	6:43 PM
20			G2	Monday, December 03, 2012	7:15 PM
21	Salah Salem Street	Backward	G1	Monday, December 17, 2012	5:45 PM
22			G2	Monday, December 17, 2012	6:16 PM
23	El-Nasr Street	Forward	G1	Wednesday, December 19, 2012	3:30 PM
24			G2	Wednesday, December 19, 2012	4:22 PM
25		Backward	G1	Wednesday, December 19, 2012	5:41 PM

26			G2	Wednesday, December 19, 2012	6:29 PM
27	Gameet El-Dewl El-Arabyia Street	Forward	G1	Thursday, January 10, 2013	10:15 AM
28			G2	Thursday, January 10, 2013	10:45 AM
29		Backward	G1	Thursday, January 10, 2013	11:13 AM
30			G2	Thursday, January 10, 2013	11:47 AM

3.4 Tool and Technique

The main technique used to collect data for preparing this research is floating cars equipped with Global Positioning System (GPS) devices. The devices were installed on both buses and automobile in order to obtain vehicle's track and use it for travel time estimation.

3.4.1 Global Positing System (GPS)

The Global Positioning System is a satellite-based navigation system made up of a network of 33 satellites placed into orbit by the U.S. Department of Defense. GPS was originally intended for military applications, but in 1980s, the government made the system available for civilian use. The GPS works in any weather conditions, anywhere in the world, 24 hours a day.

The GPS is operated by the United States government, which is solely responsible for accuracy and maintenance of the system. Location data accuracy can be affected by adjustments to GPS satellites made by the U.S. government and is subject to change with civil GPS policy and the Federal Radio navigation Plan. It can also be affected by poor satellite geometry.

3.4.2 GPS Device

In this research, the GPS device used is Columbus V-900 data logger (Figure 3.7), which is designed with a concept of "user-friendly manipulation". It is embodied by its audible and visual alerts that indicate the Bluetooth connection status, card plugging and unplugging, initiation of audio recording, over-speed alarm, mode changeover, and other operations. With these agile alerting lights and audios, a user can experience the convenience and pleasure brought about by technologies.

Figure 3.7
COLOUMBUS V-900.

The device offers many advantages such as its small size which makes it easy to carry and its ease of use, as it allows taking notes along the journey by recording voice. Moreover, a track is logged on the memory as a CSV file hence, the user can manipulate and manage the file from the computer as shown in Figure 3.8.

INDEX	TAG	DATE	TIME	LATITUDE N/S	LONGITUDE E/W	HEIGHT	SPEED	HEADING	VOX
1	T	121219	182859	30.024534N	031.257003E	74	0	0	
2	T	121219	182902	30.024530N	031.257006E	68	0	0	
3	T	121219	182903	30.024526N	031.257006E	65	0	0	
4	T	121219	182904	30.024525N	031.257004E	62	0	0	
5	T	121219	182909	30.024508N	031.257009E	59	3	0	
6	T	121219	182910	30.024501N	031.257013E	59	3	0	
7	T	121219	182911	30.024495N	031.257018E	59	4	0	
8	T	121219	182912	30.024415N	031.257116E	59	6	0	
9	T	121219	182913	30.024401N	031.257139E	59	7	0	
10	T	121219	182914	30.024428N	031.257200E	59	7	0	
11	T	121219	182915	30.024399N	031.257236E	59	8	0	
12	T	121219	182916	30.024399N	031.257260E	59	8	0	
13	T	121219	182917	30.024411N	031.257278E	59	7	0	
14	T	121219	182918	30.024423N	031.257305E	59	7	0	
15	T	121219	182919	30.024434N	031.257328E	59	6	0	
16	T	121219	182920	30.024455N	031.257346E	59	6	0	
17	T	121219	182921	30.024464N	031.257371E	59	6	0	
18	T	121219	182922	30.024473N	031.257388E	59	5	0	
19	T	121219	182923	30.024473N	031.257390E	59	4	0	
20	T	121219	182924	30.024471N	031.257419E	59	5	0	
21	T	121219	182925	30.024489N	031.257436E	59	6	0	
22	T	121219	182926	30.024495N	031.257471E	59	8	0	
23	T	121219	182927	30.024500N	031.257494E	59	7	0	
24	T	121219	182928	30.024508N	031.257511E	59	7	0	
25	T	121219	182929	30.024514N	031.257534E	59	9	0	
26	T	121219	182930	30.024523N	031.257570E	59	12	72	
27	T	121219	182931	30.024528N	031.257613E	58	13	77	
28	T	121219	182932	30.024546N	031.257648E	58	14	81	
29	T	121219	182933	30.024546N	031.257698E	58	15	82	
30	T	121219	182934	30.024551N	031.257758E	58	16	82	
31	T	121219	182935	30.024556N	031.257819E	58	18	80	

Figure 3.8 Example of a track file CSV.

3.5 Relevant Software

Having the above mentioned features; a good data logger still needs practical track processing software. “Time Album” software attached to the V-900 device helps freely processing track in addition to “Google Earth” software that is linked to the V-900 device as well.

3.5.1 Time Album

The V-900 comes with “Time Album” software that enables easy management of journey track information. With a user-friendly interface and simple operation, the software is advantageous over its general counterparts for format conversion. Journey photos and voice records can be integrated with travel tracks simply and easily, and then clearly demonstrated in a map to present a memorable, exciting journey at a glance. With excellent compatibility across multiple platforms, it can run on a PC, or on a Mac or Linux operating system. A screenshot of the software is shown in Figure 3.9.

A user may use "Time Album" to configure detailed settings of the device which at the beginning needs to set the GPS time by setting time zone of the user location, and the vehicle type: car mode or fly mode. Afterwards it allows importing the track and adding audio records and photos applicable to the current track. At the end, a final track report is created with a summary of the trip log (Figure 3.10). "Time Album" can be also used to generate a track file with an extension of KMZ to play back the journey track in Google Earth Figure 3.11.

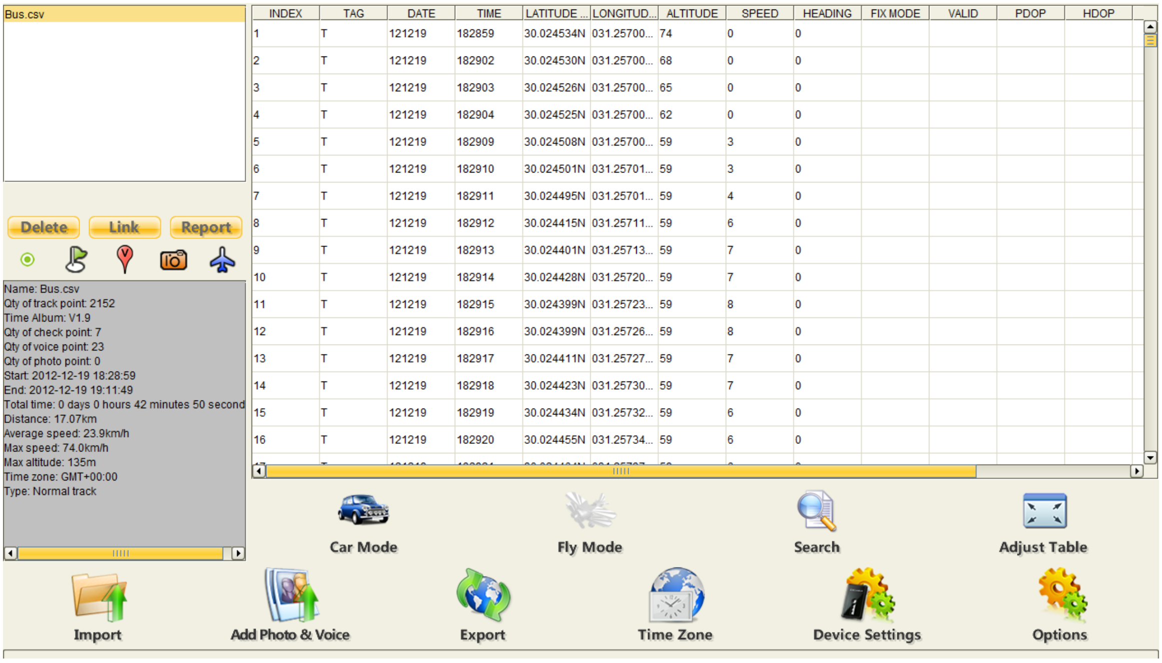

Figure 3.9 Time Album Software.